

Universidad Autónoma de Madrid



**Programa de Doctorado en Biociencias Moleculares**

---

Análisis de la diversidad genómica y transcriptómica de *Trypanosoma cruzi* y su relación con la enfermedad de Chagas.

Tesis doctoral: Francisco Callejas Hernández

Madrid, 2019

**Universidad Autónoma de Madrid**

Facultad de Ciencias

Departamento Biología Molecular

Análisis de la diversidad genómica y transcriptómica de *Trypanosoma cruzi* y su relación con la enfermedad de Chagas.

Memoria presentada por el Ingeniero en Biotecnología **Francisco Callejas Hernández**

**Directores:**

Dr. Manuel Fresno Escudero

Dra. Núria Gironès Pujol

Centro de Biología Molecular Severo Ochoa  
Madrid, 2019

## Agradecimientos

Este trabajo se realizó gracias a la beca pre-doctoral número **411595** otorgada por el Consejo Nacional de Ciencia y Tecnología (CONACYT, México), gestionada por el Consejo de Ciencia, Tecnología e Innovación de Hidalgo (CITNOVA, México).



*A mis padres, hermanos, mi compañera de aventuras Sam,  
Amigos  
y a mi México querido.*

*“It is hard to overstate the importance of DNA sequencing to biological research; at the most fundamental level it is how we measure one of the major properties by which terrestrial life forms can be defined and differentiated from each other.”*

James M. Heather and Benjamin Chain

## Resumen

El parásito *Trypanosoma cruzi* es el causante de la enfermedad de Chagas. Esta enfermedad afecta principalmente zonas rurales de América central y Sudamérica, donde el insecto vector de transmisión es endémico. Sin embargo, debido a la migración de personas infectadas esta dolencia se ha extendido a continentes no endémicos, principalmente Europa y Asia. Se estima que existen actualmente hasta 20 millones de personas infectadas y se producen hasta 14 mil muertes por año, siendo el fallo cardiaco la afectación letal más común.

No existe una cura para esta enfermedad, por lo que la Organización Mundial de la Salud la ha incluido dentro del grupo de las 18 enfermedades tropicales más desatendidas.

El parásito presenta un ciclo de vida complejo alternando entre 3 morfologías principales y condiciones adversas de crecimiento tanto en el insecto vector como en el huésped mamífero. Esta misma complejidad se conserva a nivel molecular presentando una enorme variabilidad genética con hasta cientos de cepas descritas, sorprendentemente diversas en cuanto a variabilidad en tamaño y composición de su genoma, así como un marcado comportamiento diferencial en modelos de infección *in vitro* e *in vivo*.

Por lo anterior, en este trabajo se establecieron 4 estrategias principales cuyo objetivo general es contribuir al entendimiento de esta enfermedad que abarcan, desde la biología molecular del parásito hasta modelos de infección y desarrollo de la enfermedad *in vitro* e *in vivo*.

Dichas estrategias han permitido: 1.- obtener la secuencia genómica de 3 nuevas cepas del parásito; dos ensamblajes *de novo* mediante técnicas de secuenciación de segunda y tercera generación y una corrección usando datos de RNAseq, 2.- determinar el transcriptoma completo de la forma infectiva, así como un análisis de los mecanismos de maduración del ARNm, 3.- analizar la expresión de miARNs y el transcriptoma del huésped durante la fase temprana de la infección, y establecer la ruta completa de activación del sistema inmune y 4.- realizar un análisis transcriptómico completo de corazón de ratón en estado crónico de la infección, que permitió la identificación de las principales rutas de respuesta inmune, adquisición de la inmunidad y la cepa de *T. cruzi* más adecuada en modelos de infección *in vitro*.

## Abstract

*Trypanosoma cruzi* is the parasite causative of the Chagas disease. This disease affects mainly rural areas from central and south America, where the insect vector is endemic. However, due to the migration of infected people this disease has been spread to non-endemic continents, such as Europe and Asia. It is estimated that actually exists up to 20 million infected people and that it produces up to 14 thousand deaths per year, being the cardiac failure the most lethal complication.

There is no cure for this disease, and the World Health Organization has included it in the group of the 18 most neglected tropical diseases.

The parasite has a complex life cycle, alternating between 3 main morphological stages and adverse growing conditions in the insect vector and the mammal host. This complexity is conserved to the molecular level, showing a wide genetic variability with up to hundreds of strains described, strikingly diverse as to variations in genome composition and length, as well as a marked differential behavior on *in vitro* and *in vivo* infection models.

Consequently, 4 main strategies were established in this work, which have as the main focus to contribute to the understanding of this disease, ranging from the molecular biology of the parasite up to infection models and the disease development.

Those strategies allowed us: 1.- to obtain the genome sequence of 3 new parasite strains; two assembled *de novo* by second and third-generation sequencing technologies and one corrected by RNAseq data, 2.- to determine the whole transcriptome of the infective form and the analysis of mRNA maturation mechanisms, 3.- to analyze the expression of miRNAs and the transcriptomic profile in the host during the early phase of infection, and to establish the complete immune system pathway activation, and 4.- to perform a whole transcriptomic analysis of infected heart in the chronic stage of infection, which allowed us to identify the main pathways on immune response, immunity acquisition and the most suitable strain for infection models *in vitro*.



## ÍNDICE

Lista de abreviaturas .....	1
1 INTRODUCCIÓN .....	3
1.1 La enfermedad de Chagas .....	4
1.2 Ciclo de vida del parásito <i>Trypanosoma cruzi</i> .....	6
1.3 Biología molecular de <i>Trypanosoma cruzi</i> .....	7
1.3.1 Genómica .....	10
1.3.2 Transcriptómica .....	11
1.4 Interacción parásito-huésped y procesos de internalización .....	12
1.5 Respuesta inmune .....	15
1.6 MicroARNs en la enfermedad de Chagas .....	17
1.7 Secuenciación de ADN: historia y actualidad .....	18
1.8 Términos más comunes en análisis de genómica y transcriptómica .....	20
2 OBJETIVOS .....	23
2.1 Objetivo general .....	24
2.2 Objetivos específicos .....	24
3 MATERIALES Y MÉTODOS .....	25
3.1 Cultivo del parásito .....	26
3.2 Ratones y declaración ética .....	26
3.3 Obtención de macrófagos intraperitoneales e infección <i>in vitro</i> e <i>in vivo</i> ..	27
3.4 Extracción de ARN total .....	27
3.4.1 ARN total de la infección <i>in vitro</i> a 1, 6 y 24 horas .....	27
3.4.2 ARN total de la infección <i>in vivo</i> a 6 meses post-infección .....	28
3.4.3 ARN total de tripomastigotes de la cepa Sylvio X10 .....	28
3.5 Extracción de ADN .....	28
3.6 Secuenciación, ensamblaje y análisis de los nuevos genomas de las cepas Bug2148 e Y .....	29
3.6.1 Genes de las Trans-sialidasas, secuencias teloméricas y secuencias repetitivas en los ensamblajes .....	30
3.7 Predicción genética y anotación genómica .....	30



3.8 Secuenciación y ensamblaje del transcriptoma de Sylvio X10 en su forma tripomastigote .....	30
3.9 Disponibilidad de datos de secuenciación.....	31
3.10 MiARNseq a 1,6 y 24 horas post-infección, secuenciación y análisis.....	31
3.11 Expresión diferencial de miARNs en macrófagos en respuesta a la infección por <i>T. cruzi</i> e identificación de los potenciales ARNm diana .....	32
3.12 Identificación de potenciales miARNs de <i>T. cruzi</i> en macrófagos intraperitoneales de ratón BALB/c infectados con la cepa Y.....	32
3.13 ARNseq de macrófagos intraperitoneales de ratón BALB/c infectados con <i>T. cruzi</i> , 24 horas post-infección.....	32
3.14 ARNseq de corazón de ratones BALB/c infectados con <i>T. cruzi</i> , 6 meses post-infección.....	33
3.15 Estadística y figuras.....	34
4 RESULTADOS Y DISCUSIÓN.....	35
4.1 Secuenciación de dos nuevos genomas de <i>T. cruzi</i> .....	36
4.1.1 Genómica de <i>T. cruzi</i> , secuenciación y ensamblaje .....	36
4.1.2 Predicción genética y análisis funcional .....	38
4.1.3 Actividad Trans-sialidasa: subfamilias .....	42
4.1.4 Estructura cromosómica en <i>T. cruzi</i> y diversidad genética .....	45
4.1.5 Genes de copia única en <i>T. cruzi</i> .....	51
4.2 Transcriptómica de la forma infectiva de la cepa Sylvio X10 Y refinamiento de la secuencia genómica .....	52
4.2.1 Refinamiento de la secuencia genómica, ensamblaje del transcriptoma y cuantificación .....	52
4.2.2 Análisis de enriquecimiento funcional del transcriptoma de <i>T. cruzi</i> ...	62
4.2.3 Sitios de trans-splicing y poli-adenilación .....	64
4.3 miARNs y transcriptómica de la fase temprana de la infección por <i>T. cruzi</i> .....	67
4.3.1 Expresión diferencial de miARNs de macrófagos infectados con <i>T. cruzi</i> .....	67

4.3.2 Análisis de la expresión génica de macrófagos intraperitoneales de ratón BALB/c en respuesta a la infección por <i>T. cruzi</i> a 24 h.p.i .....	76
4.3.3 Perfil transcriptómico del parásito 24 h.p.i.....	81
4.3.4 Presencia de ARNs cortos de <i>T. cruzi</i> con potencial función de miARNs .....	82
4.4 Transcriptómica de corazón de ratón BALB/c en estado crónico de la enfermedad de Chagas .....	87
4.5 Panorama general .....	96
5 CONCLUSIONES .....	98
6 BIBLIOGRAFÍA .....	101
7 ANEXOS .....	127
7.1 Apéndice A. Material suplementario .....	128
7.1.1 Figuras Suplementarias .....	128
7.2 Apéndice B. Publicaciones relacionadas .....	146

## Lista de abreviaturas

Abreviatura	Significado
APC	Célula presentadora de antígeno ( <i>Antigen Presenting Cells</i> )
BLAST	<i>Basic Local Alignment Search Tool</i>
CCC	Cardiomiopatía chagásica crónica
C-LS	Homólogos de cromosomas ( <i>Chromosome-like sequences</i> )
CRP	Proteína del complemento regulatorio ( <i>Complement Regulatory Protein</i> )
DCM	Cardiomiopatía dilatada ( <i>Dilated cardiomyopathy</i> )
DGF	<i>Dispersed Gene Family</i>
DTUs	<i>Discreet Typing Units</i>
eVes	Vesículas extracelulares de la forma epimastigote
FPKM	Fragmentos por kilobase por millón ( <i>Fragments Per Kilobase per Million</i> )
GJs	<i>Gap junctions</i>
GO	Ontología de genes ( <i>Gene Ontology</i> )
HCM	Cardiomiopatía hipertrófica ( <i>Hypertrophic cardiomyopathy</i> )
Hsp	Proteína de choque térmico ( <i>Heat shock proteins</i> )
IGV	<i>Integrative Genomics Viewer</i>
ITAMs	Inmunorreceptores de activación basados en motivos de tirosina. ( <i>Immunoreceptor tyrosine-based activation motifs</i> )
KEGG	Enciclopedia de genes y genomas Kyoto ( <i>Kyoto Encyclopedia of Genes and Genomes</i> )
Log2Fold	Razón de cambio en base 2 logarítmica ( <i>Fold of Change</i> )
MASP	Proteína de superficie asociada a mucinas ( <i>Mucin-Associated Surface Protein</i> )
mCell	RNAs cortos de célula completa ( <i>T. cruzi</i> )
MCL	Algoritmo de agrupación de Markov ( <i>Markov Clustering Algorithm</i> )
MDCS	Células mieloides supresoras ( <i>Myeloid-derived suppressor cells</i> )
MEC	Matriz extracelular
miARNs	microARNs ( <i>miRNAs</i> )
Moco	Proteína de biosíntesis cofactor de Molibdeno ( <i>Molybdenum cofactor biosynthesis protein</i> )
mVes	Vesículas extracelulares de la forma tripomastigote metacíclico
NAFLD	Enfermedad de hígado graso no alcohólico ( <i>Non-Alcoholic Fatty Liver Disease</i> )
NCBI	<i>National Center for Biotechnology Information</i>
NGS	Secuenciación de nueva generación ( <i>Next Generation Sequencing</i> )
NK	Célula asesina natural ( <i>Natural Killer</i> )
NTDs	Enfermedades tropicales desatendidas ( <i>Neglected Tropical Diseases</i> )
OMS	Organización mundial de la salud
ORF	Marcos de lectura abiertos ( <i>Open Reading Frame</i> )

Abreviatura	Significado
PABPs	Proteínas de unión a poli-A ( <i>polyA-binding proteins</i> )
PacBio	<i>Pacific Biosciences</i>
PAMPs	Patrones moleculares asociados a patógenos ( <i>Pathogen Associated Molecular Patterns</i> )
PCOA	Análisis de componentes principales ( <i>Principal Components Analysis</i> )
PFGE	Electroforésis en campo pulsado ( <i>Pulse Field Gel Electrophoresis</i> )
PIP3	Fosfatidilinositol-3,4,5-trifosfato
PRRs	Receptores reconocedores de patrones ( <i>Pattern-Recognition Receptors</i> )
PTUs	Unidades de transcripción policistrónica ( <i>Polycistronic Transcription Units</i> )
RBP	Proteínas de unión a ARN ( <i>RNA binding proteins</i> )
RDC	Profundidad de cobertura de lecturas ( <i>Read Depth Coverage</i> )
RHS	<i>Retrotransposon Hot Spot</i>
RIN	Número de integridad de ARN ( <i>RNA integrity number</i> )
RISC	Complejo de silenciamiento inducido por RNA ( <i>RNA-Induced Silencing Complex</i> )
RNI	Intermediarios de nitrógeno reactivo ( <i>Reactive Nitrogen Intermediates</i> )
ROS	Intermediarios de oxígeno reactivo ( <i>Reactive Oxygen Species</i> )
RPKM	Lecturas por kilobase por millón ( <i>Reads Per Kilobase per Million</i> )
RRM3	Motivo 3 de reconocimiento de ARN ( <i>RNA recognition motif 3</i> )
SAPA	Antígeno de fase aguda ( <i>Shed Acute-Phase Antigen</i> )
SFB	Suero fetal bovino
SL	<i>Spliced Leader</i> (Miniexón)
SMRT	Molécula sencilla en tiempo real ( <i>Single Molecule Real-Time</i> )
SNAREs	Receptores de proteínas de fijación soluble de NSF ( <i>Soluble N-ethylmaleimide sensitive factor attachment protein receptors</i> )
SNPs	Polimorfismos de nucleótido único ( <i>Single Nucleotide Polymorphisms</i> )
SRA	Archivo de lecturas de secuenciación ( <i>Sequencing Read Archive</i> )
SSR	Región de cambio de hebra ( <i>Switch-strand region</i> )
TcMUCs	Mucinas de <i>T. cruzi</i>
TCNA	Neuramidasa de <i>T. cruzi</i>
Th	Célula cooperadora T
TPM	Tránscrios por kilobase por millón ( <i>Transcripts Per Kilobase per Million</i> )
TSs	Trans-sialidasas
UCSC	<i>University of California Santa Cruz</i>
UTR	Región no traducida ( <i>Untranslated Region</i> )
WebGestalt	<i>WEB-based GENE SeT AnaLysis Toolkit</i>

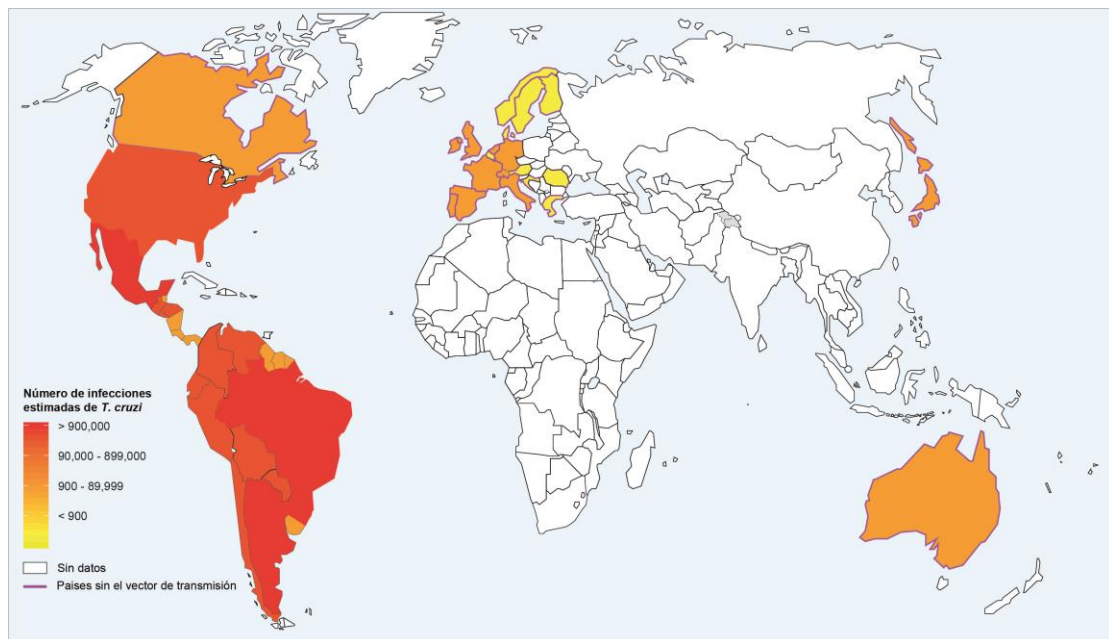
## 1 INTRODUCCIÓN

---

## 1.1 La enfermedad de Chagas

La enfermedad de Chagas, nombrada así en honor al científico Brasileño que la descubrió en 1909 (Carlos Chagas), es ocasionada por el parásito *Trypanosoma cruzi*[1]. Según datos de la organización mundial de la salud (OMS), se estima que existen actualmente 8 millones de personas con esta dolencia, 25 millones en riesgo potencial de contraerlo y produce de 10 a 14 mil muertes por año[2,3]. Sin embargo, como los datos oficiales de OMS sobre esta enfermedad no han sido actualizados desde el 2010, algunos investigadores consideran que la cifra total actual de personas infectadas podría alcanzar los 20 millones[3,4].

Es conocida también como Tripanosomiasis Americana en alusión al continente donde el vector de transmisión es endémico (a excepción de Canadá) y donde es además la enfermedad parasitaria más importante, con casos reportados en al menos 21 países. Sin embargo, esta enfermedad se ha extendido a otros continentes mediante la migración de personas infectadas, convirtiéndose en un problema de salud a nivel mundial (**Figura 1**)[2,5,6]. Además, debido a que el mayor número de casos reportados en América latina provienen de países en vías de desarrollo y zonas rurales con precarios sistemas de salud, la OMS ha incluido a esta enfermedad en el grupo de las 20 enfermedades tropicales desatendidas o NTDs (*Neglected Tropical Diseases*) y la cuarta enfermedad infecciosa más importante solo después de la malaria, tuberculosis y esquistosomiasis[2,7,8].



**Figura 1.** Casos estimados de infecciones por *T. cruzi* en todo el mundo, según datos de la OMS (2006-2010, <https://www.who.int/chagas/en/>).

El principal medio de transmisión de esta enfermedad en América latina es mediante el insecto vector, un insecto hematófago de la familia *Reduviidae* y subfamilia *Triatominae*. Existen más de 140 especies de triatominos, y se estima que al menos 40 son capaces de hospedar al parásito *T. cruzi*, siendo *Triatoma infestans* la especie más común (conocida popularmente como “*chinche besucona*”)[3,6,9].

Por otra parte, existen mecanismos de transmisión de la enfermedad no vectoriales (principalmente en zonas urbanas), los catalogados hasta el momento son: transfusiones sanguíneas (10-25% de incidencia), de manera congénita (en algunos casos dependiente de los niveles de parasitemia en la madre), trasplante de órganos o médula ósea (siendo el trasplante de corazón el más común con 75-100% de incidencia), de forma oral por la ingesta de alimentos contaminados o incluso por accidentes de laboratorio[3,6].

Clínicamente la enfermedad presenta varias fases. La fase aguda es la primera en desarrollarse y en el 70% de los casos es asintomática, o presenta síntomas leves que suelen confundirse con los de un resfriado común. Sin embargo, en los países endémicos, es posible detectar señales específicas de la infección por *T. cruzi* tras la picadura; el chagoma y el signo de Romaña nombrado así en honor al investigador Argentino que lo descubrió y describió (Cecilio Romaña)[10]. El chagoma se caracteriza como una inflamación subcutánea severa en el sitio de la picadura que suele aparecer durante los primeros 15 días, mientras que el signo de Romaña es específico de infecciones vía ocular causando inflamación severa del párpado[11].

Después de 6-8 semanas de la picadura los síntomas desaparecen llegando a la fase indeterminada, que puede durar décadas en las que la parasitemia es indetectable por métodos parasitológicos[12]. En raras ocasiones esta etapa puede presentar complicaciones y causar la muerte, siendo el fallo cardíaco en pacientes inmunodeficientes la complicación más común[7].

Después de 5-30 años de la infección inicial, aproximadamente un 30-40% de los pacientes evolucionan hacia la etapa crónica. La etapa crónica sintomática de la enfermedad se caracteriza principalmente por causar mega-vísceras (colon y esófago) y/o cardiomiopatía dilatada progresiva[7,12], siendo ésta última la principal causa de decesos por muerte súbita.

No existe cura de la enfermedad y desde la década de los 1970s, existen solamente dos fármacos aprobados para tratar esta enfermedad, Benznidazol y Nifurtimox, con tratamientos que van de los 60 hasta los 90 días cuyo objetivo principal es reducir la carga parasitaria. Ambos fármacos presentan actividad tripanocida, menor toxicidad en niños que en adultos, su eficacia es dependiente de la cepa de infección siendo

más efectivos en la etapa aguda, aunque han demostrado también cierta efectividad en la etapa crónica[13,14].

El uso de estos fármacos puede ocasionar efectos secundarios, principalmente en adultos. Estos efectos van desde náuseas, vómitos, dolor abdominal, diarrea e insomnio hasta dermatitis, alucinaciones, depresión de la médula ósea con trombocitopenia y polineuropatía[4,13,14]. Algunos investigadores recomiendan comenzar y terminar el tratamiento, excepto en personas adultas mayores de 50 años para los que se desconocen sus efectos secundarios. Sin embargo, la gran mayoría de los pacientes se ven obligados a interrumpir el tratamiento debido a los efectos secundarios.

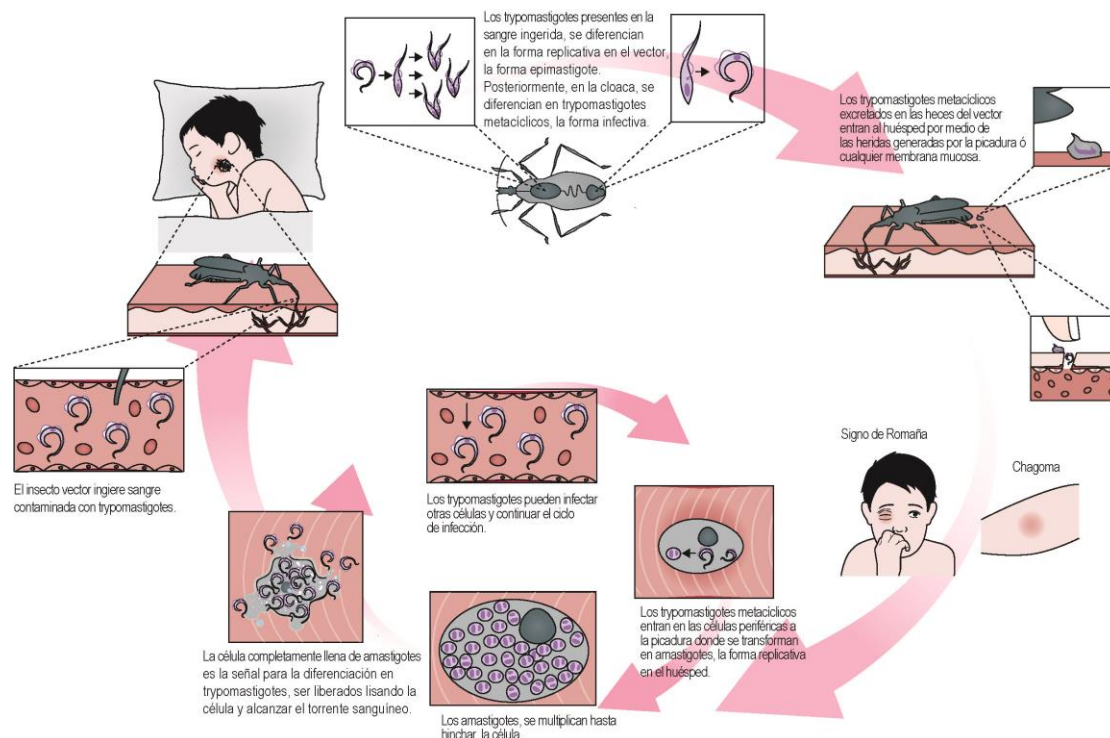
## **1.2 Ciclo de vida del parásito *Trypanosoma cruzi***

El parásito alterna entre 3 formas o etapas morfológicas principales durante su ciclo de vida, que incluye al insecto vector y el huésped mamífero. Estas formas se diferencian además de por sus características físicas, por la función que desempeñan. El trypomastigote metacíclico, es la forma flagelada y capaz de causar la infección, el amastigote es la forma redondeada no flagelada y replicativa en el huésped mamífero mientras que el epimastigote, es la forma alargada flagelada y replicativa en el insecto vector.

El ciclo inicia con la picadura del insecto triatomino infectado, principalmente mientras el huésped duerme. Mientras se alimenta defeca, y los trypomastigotes metacíclicos presentes en las heces inician el proceso de infección entrando en el huésped mediante las heridas provocadas al rascar el sitio de la picadura (**Figura 2**). Una vez dentro, los tripomastigotes metacíclicos invaden las células periféricas a la picadura y se transforman en amastigotes. Cuando la célula se encuentra completamente llena de amastigotes, el parásito se transforma nuevamente en la forma trypomastigote recuperando el flagelo y lisando la célula. Estos tripomastigotes son capaces de infectar tejidos adyacentes y propagarse por el torrente sanguíneo a otros tejidos continuando con el proceso de infección y su reproducción.

El ciclo se completa cuando el insecto pica al huésped mamífero infectado y se alimenta de su sangre contaminada por tripomastigotes. Los tripomastigotes recién ingeridos se diferencian en epimastigotes en el intestino medio del triatomino y se multiplican, posteriormente, en la cloaca son transformados nuevamente en tripomastigotes metacíclicos, quedando listos para continuar el ciclo.





**Figura 2.** Ciclo de vida de *Trypanosoma cruzi*. Adaptada de: José A Pérez-Molina e Israel Molina (2017), página 2, Figura 1[3].

Dada la importancia del insecto vector en el ciclo de vida del parásito y en el ciclo de infección, en los últimos años se han propuesto diversas iniciativas en contra de esta enfermedad, cuyo propósito principal es su erradicación y se ha demostrado que es el método más efectivo para prevenir la infección[15]. Estas iniciativas incluyen, además, test a donantes de sangre, mujeres embarazadas, detección temprana y tratamientos oportunos.

La participación activa en estas iniciativas de los países endémicos del vector de transmisión ha propiciado grandes avances, principalmente en países como Uruguay, Chile, Brasil, El Salvador, Argentina, Bolivia y en algunas zonas de América central que han logrado el control de este insecto, reduciendo drásticamente los niveles de infección mediante esta vía[16]. Además, el control de infección por transfusiones sanguíneas ha caído de un 7 a 0.6%[17].

### 1.3 Biología molecular de *Trypanosoma cruzi*

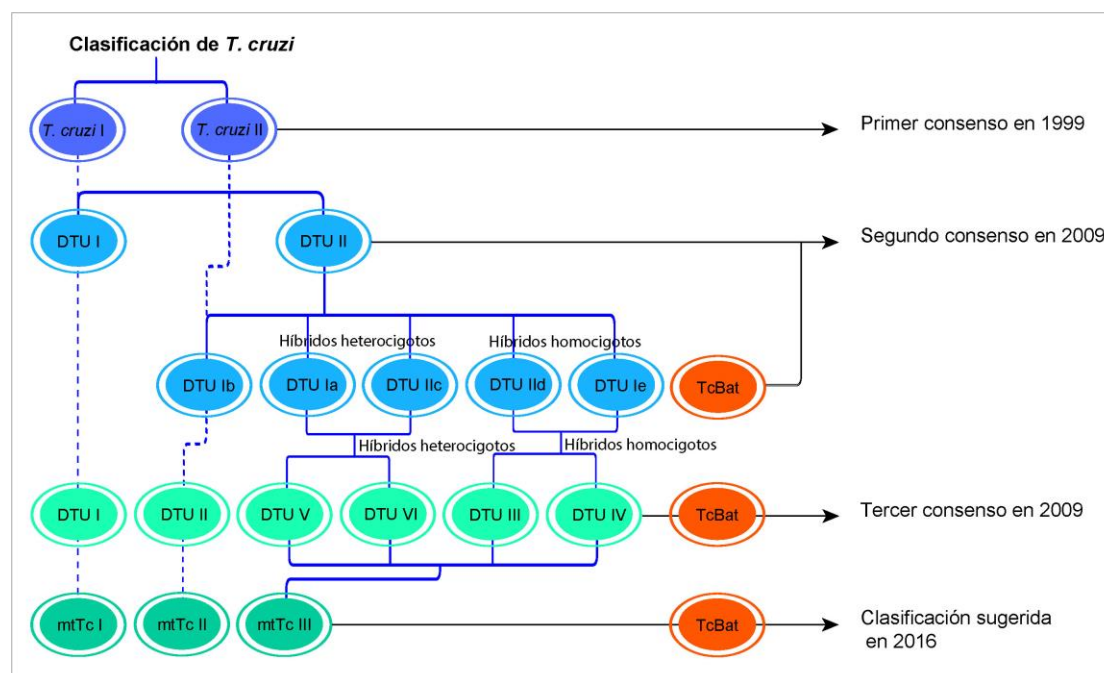
La complejidad del genoma de *T. cruzi* es un reflejo de la complejidad de su ciclo de vida. Es un organismo capaz de infectar casi cualquier célula nucleada de cualquier tejido, capaz de sobrevivir en el tracto digestivo de un insecto y capaz de sobrevivir después de haber sido eliminado en forma de heces.

Como se ha descrito anteriormente, la reproducción de este parásito es normalmente asexual por división binaria, sin embargo, existe evidencia de hibridación natural, intercambio genético entre cepas distintas y reproducción sexual que hace replantearse esta afirmación, pero que además, parecen ser las causas de un genoma especialmente complejo[18–20].

En *T. cruzi*, la mitosis ocurre sin que la membrana nuclear se fragmente completamente y aunque contiene nucleosomas, el contenido genómico no condensa hasta formar cromosomas impidiendo ser visualizados como en cualquier eucariota con técnicas convencionales[21,22]. A pesar de que existen algunos intentos, actualmente el cariotipo genómico no se ha conseguido. Sin embargo, diversos estudios de electroforesis en campo pulsado o PFGE por sus en inglés (*Pulse Field Gel Electrophoresis*) han demostrado la existencia de una gran variabilidad molecular, tanto en número de cromosomas como en su tamaño, entre diferentes cepas, y sorprendentemente incluso entre diferentes clones de la misma cepa[23,24]. Estos estudios han permitido además concluir que el parásito es principalmente diploide y que el tamaño y número de sus cromosomas es variable, desde 0.45 hasta 4 Mb de longitud y desde 19 hasta 40 cromosomas por célula[25]. Técnicas alternativas como la citometría de flujo han confirmado la variabilidad genética de este microorganismo y han permitido estimar el tamaño total del genoma, que va desde los 80 hasta los 150 Mb entre las cepas contempladas en el análisis[26]. Hasta hoy no se ha confirmado ninguna variable directa que determine esta enorme variabilidad genómica, inusual dentro de una misma especie, aunque se han propuesto algunos factores externos adicionales al complejo ciclo de vida tales como, localización geográfica y fuente de aislamiento. Algunos autores incluso plantean la posibilidad de que *T. cruzi* no sea una sola especie, sino un grupo de sub-especies relacionadas[25].

Hasta hoy se han descrito al menos 50 cepas representativas diferentes (sin contabilizar clones)[27]. En 1999, con las cepas descritas y conocimientos disponibles se llegó a un consenso de clasificación basado en características biológicas, bioquímicas y moleculares, que consideraba solamente dos grupos nombrados como *T. cruzi I* y *T. cruzi II*. Diez años después, tras un genotipado multi-*locus* más amplio, se propone la clasificación basada en 6 DTUs (*Discreet Typing Units*) agrupadas en dos grupos principales llamados DTU I y DTU II, donde la DTU II contenía las subdivisiones IIa a IIe. Se consideraba a las DTU I y DTU IIb como las cepas ancestrales, mientras que las DTUs IIc y IIe eran cepas con al menos un evento de recombinación, las DTUs IIa y IIc eran consideradas como ancestrales híbridas (**Figura 3**). En una segunda reunión científica ese mismo año (2009), se

propone un segundo consenso de clasificación en 6 DTUs (I-VI), donde las DTUs I y II seguirían clasificando a las cepas ancestrales, las DTUs III-IV a cepas con al menos un evento de recombinación entre cepas de las DTUs I y II (híbridos homocigotos), y las DTUs V y VI contendrían a las cepas híbridas entre cepas de las DTUs II y III (híbridos heterocigotos). La cepa CI Brener fue la reclasificación más notable, pasando de la DTU II a la VI[27]. Meses después se incluye una séptima DTU que clasifica a cepas aisladas en murciélagos (TcBat)[28].



**Figura 3.** Clasificaciones propuestas de las distintas cepas conocidas de *Trypanosoma cruzi* basadas en diversos marcadores moleculares, mitocondriales y nucleares.

Más tarde en 2016 Bernabé y col.[29] cuestiona este consenso de clasificación destacando como principales motivos, la ausencia de validez estadística, el número de cepas analizadas y las regiones geográficas contempladas. Subraya, además que a pesar de la gran aceptación a nivel mundial del consenso propuesto no existe un estudio de filogenético válido (estadísticamente) que muestre la clasificación de los 7 grupos, por lo tanto, propone la clasificación de las cepas en 3 grupos principales basados en el análisis filogenético de 3 genes, dos mitocondriales (*Cyt* y *COII*) y un nuclear (*Gpi*). Estos 3 grupos nombrados mtTcI-III, confirman la existencia de las cepas ancestrales inicialmente nombradas como *T. cruzi* I y *T. cruzi* II (ahora mtTcI y mtTcII) y la mezcla de cepas híbridas (mtTcIII). Los autores enfatizan en que no descartan la existencia de las 6 DTUs, sino que simplemente el

modelo de clasificación no tiene respaldo estadístico. Tampoco descarta la existencia de la TcBat como una DTU independiente, aunque se encuentre filogenéticamente relacionada con la mtTcl, sin embargo, advierte que el nombre podría ser inadecuado debido a que no es la única DTU presente en murciélagos y a que también es capaz de infectar humanos.

### 1.3.1 Genómica

La primera versión del genoma de una cepa de *T. cruzi* se publicó en 2005[30]. Se eligió a la cepa CI Brener como cepa representativa por ser una de las cepas más conocidas y utilizadas en investigación, con resultados reproducibles en modelos *in vitro*, capaz de desarrollar fase aguda sintomática y susceptibilidad a Benznidazol[31]. Sin embargo, mientras que otros kinetoplastidos cercanos tales como *Leishmania* y *T. brucei* contienen genomas relativamente complejos con repeticiones que se corresponden entre el 20 y 25% de la secuencia total, en *T. cruzi* el 50% de su genoma se corresponde a este tipo de secuencias[32], complicando de manera inusual el ensamblaje de su genoma completo. La mayoría de repeticiones identificadas consisten en familias multi-génicas, repeticiones en tándem, repeticiones cortas, largas y elementos retrotransponibles[31,32]. Debido a que actualmente no existe ninguna tecnología capaz de secuenciar cromosomas enteros, y que CI Brener es una cepa híbrida, donde sus cromosomas homólogos presentan grandes diferencias de longitud y contenido genético, la primera versión del genoma estuvo lejos de haber sido reconstruido en su totalidad y con la calidad esperada[33]. El resultado fue un genoma incompleto y fragmentado que limitó su principal objetivo; constituir una referencia fiable. No obstante, sigue siendo considerado como tal en un gran número de investigaciones.

Estudios posteriores confirmaron la complejidad del genoma a distintos niveles, diferencias en longitud de sus cromosomas entre cepas de diferente DTU, de la misma DTU, y entre clones de la misma cepa[22,34]. Se han descrito además aneuploidias atípicas como monosomías, trisomías y hasta tetrasomías, principalmente en cepas clasificadas como DTU II[22].

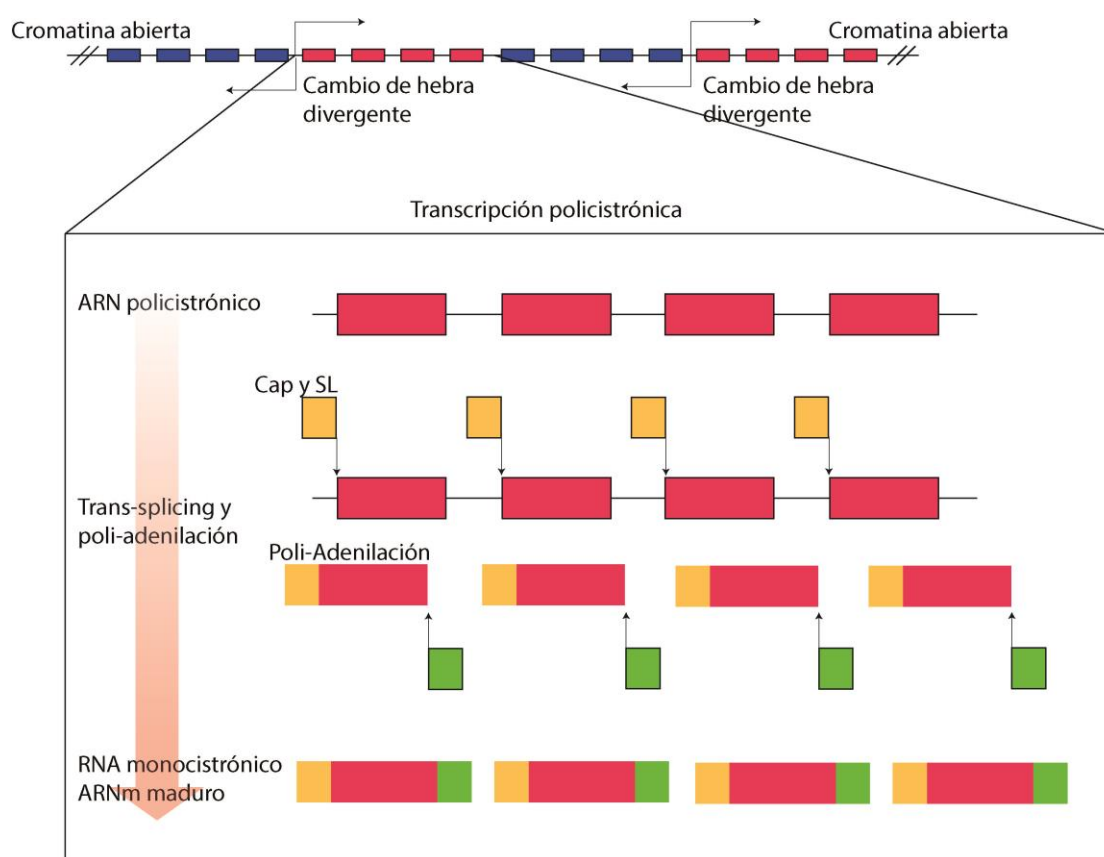
La causa de la plasticidad genómica a nivel secuencia en *T. cruzi* se atribuye a la composición genética[22,25]. Se han descrito dos grupos principales; genes altamente conservados y genes de evolución constante (en ocasiones llamados *core* y *disruptive*, respectivamente). Dentro de los genes de evolución constante se encuentran principalmente genes únicos en *T. cruzi* que codifican proteínas de superficie, tales como trans-sialidasas, mucinas, y proteínas de superficie asociadas

a mucinas o MASP (*Mucin-Associated Surface Protein*). Por otro lado, en el grupo de genes altamente conservados se encuentran principalmente genes hipotéticos y genes conservados en otras especies relacionadas como *Leishmania* y *T. brucei*[32].

### 1.3.2 Transcriptómica

Contrario a la mayoría de los organismos eucariotas, la transcripción genética en *T. cruzi* ocurre de manera policistrónica, y a diferencia de los operones bacterianos donde este mecanismo ha sido ampliamente estudiado, los genes contenidos necesitan de procesos de maduración adicionales para ser traducidos[35,36].

La transcripción inicia en cambios de sentido divergentes de la hebra codificante y termina en regiones convergentes[37]. Grupos largos de genes sin relación funcional son transcritos al mismo tiempo por la RNAPol-II en lo que se conoce como unidades transcripcionales policistrónicas ó PTUs (*Polycistronic Transcription Units*)[36]. Posteriormente, mediante procesos de trans-splicing y poli-adenilación se generan ARNm individuales maduros y estables (**Figura 4**).



**Figura 4.** Expresión génica en *T. cruzi*. Maduración de ARNm por trans-splicing y poli-adenilación.

El trans-splicing es parte del proceso de maduración del ARN policistrónico, consiste en la inserción de una secuencia de 39 nucleótidos en el extremo 5' de cada transcrito, conocido como secuencia mini-exón o SL (*Spliced Leader*) que brinda estabilidad al ARNm, y que causa la escisión de los ARNm codificantes individuales del ARN policistrónico, permitiendo la adición del poli-A. Por otro lado, la poliadenilación es un proceso de maduración común al resto de eucariotas, que consiste en insertar una cadena o cola de adeninas de longitud variable en el extremo 3' de cada transcrito (**Figura 4**)[35,36].

En ausencia de promotores que regulen la transcripción, la estabilización, desestabilización y la eficiencia en la traducción de los transcritos son los principales mecanismos que controlan la expresión génica en tripanosomátidos, además de la regulación por histonas modificadas[36,38]. La transcripción policistrónica implica que todos los genes contenidos en una misma PTU son transcritos en la misma proporción, sin embargo, se ha descrito que genes adyacentes de la misma PTU muestran distintos niveles de abundancia relativa en las diferentes etapas del ciclo de vida del parásito, o en diferentes condiciones de crecimiento[39–41].

Mientras que en otros parásitos relacionados con *T. cruzi*, como *Leishmania* y *T. brucei* los mecanismos moleculares relacionados con el proceso de expresión génica han sido ampliamente estudiados, para esta especie existe muy poca información[42].

#### 1.4 Interacción parásito-huésped y procesos de internalización

El parásito es capaz de infectar diversos tipos celulares, principalmente macrófagos, fibroblastos y células del epitelio mediante un proceso de reconocimiento e interacción mutuo que incluye gran variedad de moléculas, tanto del parásito como de la célula huésped (**Figura 5**)[43]. Las diferentes cepas conocidas, así como las formas infectivas del parásito expresan diferentes moléculas en su superficie, por lo tanto, los mecanismos de interacción e invasión son dependientes del tipo celular y de la cepa[44]. Además, se ha descrito en modelos de infección *in vivo* la influencia del fondo genético en los procesos de interacción, internalización y respuesta inmune, siendo en la mayoría de los casos la cepa BALB/c la cepa más susceptible y C57BL/6 la cepa más resistente[45,46].

El proceso infección se desarrolla en 3 etapas principales, adhesión-reconocimiento, señalización e invasión, involucrando varios procesos biológicos, desde la formación de una vacuola de endocitosis (conocida como vacuola parasitaria o VP) que

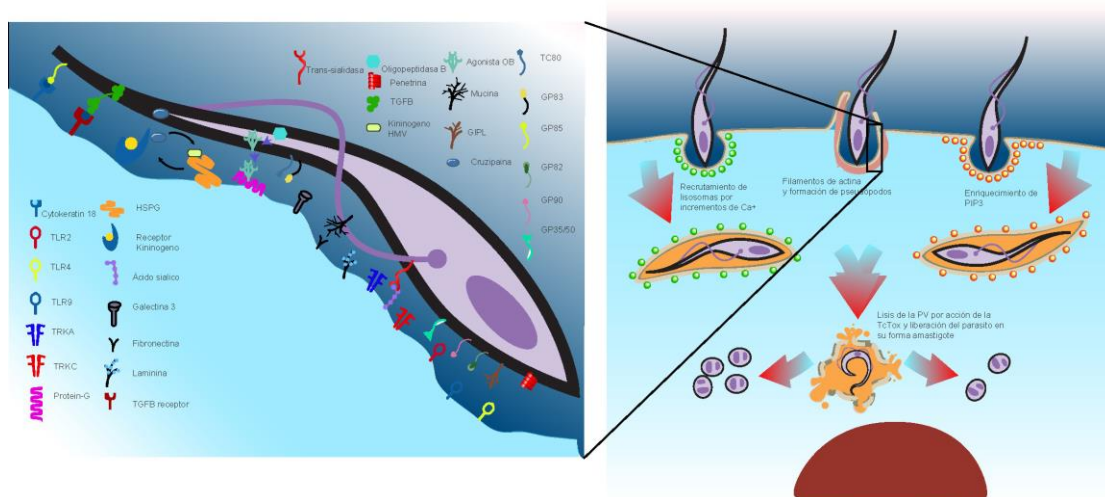
internaliza al parásito, el cambio en la forma del parásito (a la forma amastigote), la lisis de la vacuola de internalización y el contacto directo del parásito con los orgánulos de la célula huésped[47].

Las principales moléculas del parásito involucradas en este proceso de interacción con el huésped se componen principalmente por glicoproteínas de superficie, tales como la *gp90*, *Tc85*, *gp85*, *gp82*, *gp83*, *gp35/50*, mucinas, MASP y otras proteínas de superficie como las trans-sialidasas, oligopeptidasa B y cruzipaina. Algunas de estas proteínas como la *Tc85*, *gp82*, *gp35/50* y *gp83* son específicas de la forma tripomastigote[43,48,49] y sus funciones son variadas, desde la actividad glicosidasa hasta actuar como simples ligandos.

Las trans-sialidasas (TSs) de *T. cruzi* son proteínas únicas de la especie, y a nivel funcional son una modificación de las sialidasas en otras especies, su función principal es la transferencia de ácido siálico en la membrana de la célula huésped al parásito que no es capaz de producirlo[50], sin embargo, se ha demostrado también su importancia en la estabilización de la interacción[51]. Las TSs son además una de las familias de proteínas más abundantes en *T. cruzi*, que por su variabilidad puede clasificarse en 8 grupos principales, donde cada grupo contiene una estructura y función específicos[52]. En 1996 un estudio de Pereira y col.[53] usando parásitos TS<sup>-</sup> y TS<sup>+</sup>, demostraron que las poblaciones TS<sup>+</sup> tienen un alta capacidad de infección comparados con la población TS<sup>-</sup>.

Otra de las familias genéticas más abundantes de *T. cruzi* que han sido implicadas en el proceso de infección son las mucinas (conocidas como TcMUCs) y las MASP (similares en estructura, pero no en secuencia a las TcMucs). Las dos principales funciones de esta extensa familia son la protección del parásito del entorno hostil durante su ciclo de vida y la especificidad de unión a células y tejidos[51,54]. Se ha demostrado que los residuos de carbohidratos presentes en estas proteínas interactúan de forma selectiva con diferentes receptores de la célula huésped, principalmente como ligandos[43]. Además, debido a la gran variabilidad a nivel de secuencia que presentan, constituyen un factor importante en la habilidad del parásito para infectar múltiples tipos celulares así como en la evasión del sistema inmune[51].

Por otra parte, las moléculas del huésped mamífero involucradas en la interacción con el parásito son igualmente abundantes y dependientes del tipo celular, por lo que es considerado un proceso multifactorial (**Figura 5; izquierda**).



**Figura 5.** Interacción parásito-huésped y procesos principales de internalización. Adaptada de De Souza y col, 2010, páginas 4 y 13[43].

La citoqueratina 18 una integrina del cito-esqueleto encargada de la interacción célula-célula o célula-matriz extracelular, es uno de los principales receptores del parásito mediante el reconocimiento de la Tc85 y capaz de producir la activación de la ruta PI-3 cinasa. El factor- $\beta$  de crecimiento transformante o *Tgf- $\beta$*  (*Transforming growth factor- $\beta$* ) es otra molécula de superficie capaz de reconocer al parásito principalmente en células epiteliales, sin embargo, se desconoce su potencial ligando. El receptor 2 de bradiquinina, y los receptores de factores de crecimiento nervioso *TrkA* y *TrkC* (presentes en células neuronales y dendríticas) a través de la cruzipaina y TSs, respectivamente, son otras clases de receptores que influyen en el proceso de internalización del parásito[43]. Finalmente, la galectina-3 es una proteína de unión a carbohidratos relacionada con la unión específica a patógenos, indispensable en el proceso de interacción y desarrollo de la respuesta inmune[55]. Los receptores más ampliamente estudiados en modelos de ratón *in vivo* e *in vitro* son los receptores reconocedores de patrones o PRRs (*Pattern-Recognition Receptors*) entre los que destacan los de la familia *Toll-like* o TLRs 2,4,7 y 9[43,44]. Los TLRs desempeñan un rol fundamental en la resistencia a la infección por *T. cruzi* mediante la activación de diversas rutas del sistema inmune. Los ligandos actualmente descritos de estos receptores de membrana (TLRs 2 y 4) son principalmente las glicoproteínas *gp90*, *gp35/50*, *gp83*, *gp82* y *Gipl*, mientras que para los receptores internos (*Tlr7* y *Tlr9*) el reconocimiento de moléculas de RNA y DNA, respectivamente desencadena la activación celular y la producción de citoquinas[44].



Una vez que los procesos de interacción huésped-parásito ha tenido lugar, la invasión se completa mediante una serie de procesos de señalización dependientes de incrementos transitorios de  $\text{Ca}^{2+}$  (**Figura 5, derecha**), tanto en el parásito como en el huésped[43]. Los principales mecanismos descritos de la entrada de *T. cruzi* en células fagocíticas son 3[43]. El primero es conocido como fagocitosis, mediada por la formación de pseudópodos, la activación de la ruta PI-3 cinasa y la polimerización de filamentos de actina. El segundo mecanismo de entrada consiste en endocitosis con la participación de filamentos de actina, pero sin la formación de pseudópodos, y el tercer mecanismo es dependiente de la formación de invaginaciones de la membrana, sin la formación de pseudópodos y sin la intervención de filamentos de actina, siendo el único proceso de internalización activo y por lo tanto significa un gasto de energía.

Por otra parte, mecanismos dependientes e independientes de lisosomas constituyen las dos principales vías de internalización del parásito en células no fagocíticas[56]. En el mecanismo dependiente de lisosomas (que ocurre solamente en el 20% de los casos), los incrementos de  $\text{Ca}^{2+}$  regulan la producción y exocitosis de lisosomas en la membrana del huésped en las zonas de interacción con el parásito. En contraparte, el mecanismo independiente de lisosomas consiste en la internalización mediante invaginaciones de la membrana enriquecidas con la fosfatidilinositol-3,4,5-trifosfato (*Pip3*), como producto de la activación de la ruta PI-3 cinasa[56].

Independientemente del mecanismo de internalización, el parásito es alojado en una VP de la que se libera rápidamente con la expresión de TSs/neuraminidasas que transfieren el ácido sialico presente en la membrana plasmática de la célula huésped a la suya propia, causando que la membrana de la VP sea susceptible a variaciones de pH causados por la acción de la TcTox, que finalmente la destruye liberando al parásito en el citoplasma[43,56].

## 1.5 Respuesta inmune

La respuesta inmunitaria del huésped mamífero se caracteriza por al menos 3 mecanismos principales, la eliminación del parásito (mediante macrófagos y células dendríticas), la activación de células presentadoras de antígeno o APC (*Antigen-Presenting Cells*) y la detección de células infectadas por parte de los linfocitos CD8+, así como cooperadores en el contexto del complejo principal de histocompatibilidad (MHC)[57].

En los primeros 7-10 días después de la infección, la respuesta inmunitaria innata regula los niveles de parásito mediante la producción de agentes anti-microbianos, principalmente especies reactivas de oxígeno o ROS por sus siglas en inglés (*Reactive Oxygen Species*) e intermediarios reactivos de nitrógeno o RNI (*Reactive Nitrogen Intermediates*)[58,59]. A su vez, la producción de ROS y RNI en macrófagos células NK y células T $\gamma\delta$  aumentan los niveles de diferentes citoquinas, principalmente *Il-12*, *Tnfa* e *Ifn- $\gamma$* [44].

Posteriormente, con la inmunidad adquirida mediante los linfocitos T cooperadores tipo 1, o Th1 (*T-helper 1*) y la producción de IFN $\gamma$ , se activa la producción de células fagocíticas cuyo objetivo es reducir los niveles de parásitos. Se ha descrito además el rol fundamental de las células Th1 en la producción de anticuerpos y células T citotóxicas CD8<sup>+</sup>, su ausencia o bloqueo molecular produce niveles incontrolados de parásito y la muerte[44]. Además, la respuesta inmune mediada por la activación de células B con la posterior producción de anticuerpos que circulan en el torrente sanguíneo capaces de unirse de forma específica a los antígenos de la infección, es un proceso fundamental en la producción de células B de memoria y los procesos de resistencia a la infección[60].

La activación de células Th1 se caracteriza por niveles elevados de IFN- $\gamma$ , IL-2, IL-12 y TNF- $\beta$ , mientras que Th2 producen IL-4, IL-5, IL-6, IL-9, IL-10 e IL-13[61].

Se ha demostrado además que existe una alta correlación entre la severidad de la fase aguda y los niveles de parasitemia, mientras que la fase crónica es caracterizada por la baja concentración de parásitos pero con mayor daño en los tejidos (principalmente cardíaco) que suele ser letal sin tratamiento[59,62].

La elevada carga parasitaria en la fase aguda es reducida por el sistema inmunitario hasta bajos o indetectables niveles en la fase crónica. La cardiomiopatía chagásica crónica (CCC) es una de las afectaciones más graves en estado crónico de la enfermedad, y es descrita como una cardiomiopatía inflamatoria que se desarrolla en un 30-40% de los casos después de 5-30 años de la infección[58,63,64]. Aproximadamente 10% de los pacientes con CCC desarrollan alguna de las formas letales, como disfunción ventricular, fallo cardíaco y arritmias[63].

La mayoría investigadores considera y coincide en que la afectación desarrollada por la CCC es generada por el daño acumulado e irreparable del tejido cardíaco, sin embargo, las causas específicas de su origen sigue siendo debatidas[63]. Las dos principales teorías del origen de la CCC son, la persistencia del parásito y la autoinmunidad adquirida después de la elevada expresión de citoquinas[65]. Aunque se ha demostrado la presencia de células T auto-reactivas capaces de causar daño

cardiaco, sin embargo, la contribución de esta auto-reactividad en el desarrollo y afectaciones de la CCC en humanos no está claramente definida[65–67].

## 1.6 MicroARNs en la enfermedad de Chagas

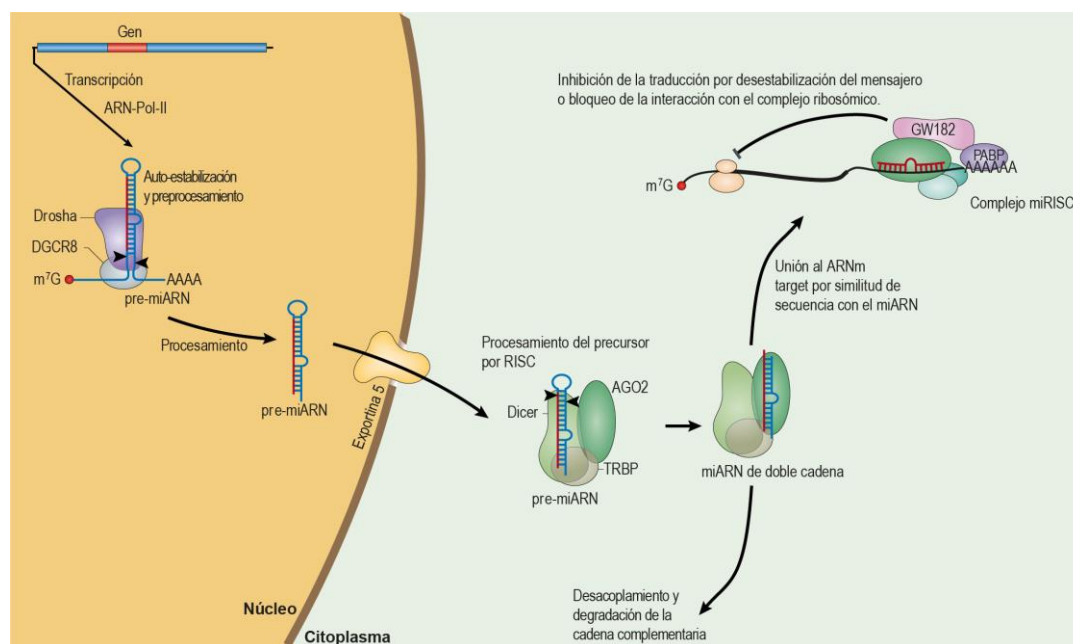
Otra clase de moléculas implicadas en el desarrollo de la enfermedad de Chagas son los microARNs (miARNs). Los miARNs son un tipo de ARN endógeno, de cadena sencilla y corta longitud (17-23 nucleótidos) capaces de reprimir de forma directa o indirecta la expresión de hasta cientos de genes, afectando la producción de proteínas por la inhibición de la traducción o la desestabilización del ARNm[68]. En humanos, se calcula que hasta el 60% de los genes codificantes son regulados por la acción de estos miARNs[69]. Además, al día de hoy se ha comprobado que los miARN participan en complejas rutas reguladoras de la expresión génica y que, variaciones en su expresión pueden desencadenar diversas patologías, como el cáncer[68].

Los miARNs pueden generarse a partir de diversos tipos de ARN, sin embargo, la biogénesis canónica consiste en la transcripción de un gen miARN o mir-gen largo por la ARN-polimerasa II, conocido como pre-miARN (**Figura 6**)[70]. Este pre-miARN se auto-estabiliza por la formación de una estructura de doble cadena y es procesado por el complejo formado por la RNA-asa III *Drosha* y la proteína de unión a ARN *Dgcr8* generando un precursor más corto de alrededor de 70 nucleótidos que es exportado al citoplasma por la exportina-5[71,72]. Posteriormente, este precursor es anclado al mecanismo de procesamiento y silenciamiento de ARN o RISC (*RNA-Induced Silencing Complex*) de forma selectiva por la acción de la proteína Argonauta 2, es procesado por el *Dicer* a la longitud final del miRNA y se libera la cadena complementaria del pre-miARN[73,74]. Finalmente, el miARN maduro es alineado parcialmente por complementariedad de bases al ARNm diana, causando su degradación por desestabilización en el extremo 3' o impidiendo su expresión por el reclutamiento de la maquinaria ribosomal de traducción[69].

Aunque los miARNs se han descrito en gran variedad de especies, el mecanismo de RISC, se ha perdido en algunos organismos, principalmente parásitos como *T. cruzi* donde no ha sido posible identificar genes codificantes a las proteínas del tipo *Argonauta* y *Dicer*, principalmente[75].

Por otra parte, esta clase de ARNs cortos tienen un rol fundamental en la fisiología y fisiopatología del sistema cardiovascular humano y murino en respuesta a la infección por *T. cruzi*, específicamente con la alteración en los niveles de los

miARNs *miR-1*, *miR-208a*, *miR-208b*, *miR-133b* y *miR-133a-2* en pacientes con CCC. Los ARNm diana de estos miARNs han sido implicados en fallo cardiaco, hipertrofia en miocardio y fibrosis[76].



**Figura 6.** Biogénesis de miRNAs (ruta canónica) en células eucariotas.

## 1.7 Secuenciación de ADN: historia y actualidad

El descubrimiento de la estructura doble hélice del ADN así como los nucleótidos que la conforman (1953)[77], inició una nueva etapa en la biología, enmarcada por la importancia de la decodificación de las secuencias genómicas de los organismos vivos, así como para entender las enfermedades relacionadas con la información genética que contienen[78].

La primera técnica de secuenciación fue la desarrollada en 1977 por Sanger y col.[79], que junto con el método desarrollado por Alan Maxam y Walter Gilbert[80] (Método Maxam-Gilbert) constituyeron la secuenciación de primera generación. La técnica de Sanger fue por décadas la tecnología de secuenciación más empleada por su alta eficiencia y baja radioactividad, comercializada como la “Tecnología de secuenciación Sanger” que perdura hasta estos días con el mismo principio científico. Se caracteriza por ser capaz de generar lecturas de poco menos de 1 Kb, de alta calidad, pero necesitando de procesos moleculares complementarios y lentos, tales como la clonación y purificación de los fragmentos a secuenciar. Entre los proyectos más importantes de esta técnica destacan principalmente, el genoma del bacteriófago “PhiX”, el primer genoma de ADN secuenciado, que por su calidad de secuencia se continúa usando como control positivo de secuenciación y

ensamblaje *in silico*[79,81], y el genoma humano, que tardó más de 15 años en ser completado, la colaboración de varios laboratorios de todo el mundo y un costo total aproximado de 100 millones de dólares[78].

Años más tarde, una nueva era de tecnologías y servicios de secuenciación emergió en 2005[82] (pirosecuenciación más tarde Roche 454, Ion torrent, Illumina y ABI/Solid) reemplazando rápidamente a las anteriores. Estas nuevas tecnologías conocidas como “secuenciación de nueva generación” o NGS (*Next Generation Sequencing*) también llamadas secuenciación de alto rendimiento (*High throughput sequencing*) se caracterizan por su capacidad de paralelización de análisis múltiples (obteniendo hasta varios millones de secuencias en unas cuantas horas), la gran cantidad de datos que generan y su bajo costo[78,81,83,84].

La mayoría de las tecnologías de secuenciación de primera y segunda generación se basan en la secuenciación por síntesis (excepto ABI SOLiD, basada en ligación y complementariedad de bases) midiendo la fluorescencia de los nucleótidos incorporados. Pacific Biosciences (PacBio), tecnología de tercera generación se basa en el mismo principio de síntesis y medición de la fluorescencia, sin embargo, solo se secuencia 1 molécula a la vez, sin pasos de amplificación (SMRT; *Single Molecule Real-Time*) permitiendo obtener lecturas de mucha mayor longitud. Por su parte *Oxford nanopore* creadora de las más novedosas y recientes tecnologías de secuenciación, utiliza la variación en la corriente eléctrica específica para cada nucleótido, a través de nanoporos artificiales por el paso de la molécula de ADN[81]. Las principales diferencias entre estas tecnologías de segunda y tercera generación se resumen en la cantidad de lecturas que producen, su longitud, así como el tipo y cantidad de errores que generan (**Tabla 1**). Los principales tipos errores generados por la tecnología Illumina son las sustituciones puntuales (*mismatches*)[85] pero contiene el menor porcentaje de error y es capaz de producir la mayor cantidad de información (hasta 6 billones de secuencias en unos días) entre las tecnologías más usadas actualmente (Illumina y PacBio, principalmente). Mientras que las tecnologías de secuenciación de tercera generación son capaces de producir las lecturas más largas hasta el momento, pero con el porcentaje de error más alto de todas las tecnologías (hasta el 17%), siendo necesario elevar drásticamente la cantidad de lecturas necesarias para obtener una secuencia consenso de calidad, o recurrir al uso de una segunda tecnología que por cobertura corrija los errores generados[86] (proceso conocido como ensamblaje híbrido), elevando así los costes de secuenciación. Los errores más comunes en estas tecnologías consisten principalmente en inserciones, deleciones y sustituciones[78,83].

Actualmente no existe ninguna tecnología en el mercado capaz de secuenciar RNA directamente, sin embargo, la tecnología Illumina ofrece la posibilidad de secuenciar RNA (total o ARNm) tras un proceso de síntesis de ADNc, permitiendo secuenciar transcriptomas completos y de alta calidad en unas cuantas horas.

**Tabla 1.** Resumen de las principales tecnologías de secuenciación de ADN de primera, segunda y tercera generación.

Generación	Tecnología	Lecturas/Carrera	Longitud de lecturas	% de Error	Lanzamiento*
Primera	Sanger	96	400-900 (pb)	0.3	1977
Segunda	Roche 454	100 - 1M**	100-700 (pb)	1	2005
	Illumina	25M - 6B***	50 - 300 (pb)	0.1 - 1	2011
	ABI SOLiD	3B - 6B	75 (pb)	0.1	2011
	Ion Torrent	400,000 - 80M	200-400	1	2011
Tercera	Pacific Biosciences	55,000	20 kb (1D) - 40 kb (2D)	15	2011
	Oxford Nanopore	2M - 4M	>200 kb	17	2014

\*Primera versión de la tecnología, \*\*Millones de secuencias, \*\*\*Billones de secuencias.

## 1.8 Términos más comunes en análisis de genómica y transcriptómica

Aunque el desarrollo de software de interfaz gráfica, amigable con personas sin formación especializada en el análisis de datos de NGS ha evolucionado de la misma forma que las tecnologías de secuenciación[87,88], algunos análisis bioinformáticos más profundos han causado que los investigadores pasen del laboratorio a la computadora, cambiando la electroforesis por algunas líneas de código, haciendo necesaria la introducción a algunos términos atípicos de la biología molecular clásica.

Dentro de los términos referentes a análisis genómicos más comunes y principalmente los empleados en este trabajo, destacan:

**Ensamblaje *de novo*:** Independientemente de la tecnología empleada y del tipo de análisis (genómica o transcriptómica), el ensamblaje *de novo* hace referencia al ensamblaje o determinación de una secuencia nucleotídica consenso a partir de las lecturas de secuenciación, sin ningún tipo de referencia que garantice la existencia del resultado por homología con resultados previos[89]. Como un rompecabezas formado por miles de piezas, con dos opciones para ser completado, la primera sin ninguna ayuda más que la similitud de piezas próximas (*de novo*), y la segunda opción con la imagen del resultado final como fondo del tablero. Sin embargo, a

diferencia de un rompecabezas, los ensamblajes *de novo* no son capaces de reconstruir la referencia original en su totalidad (sin la cobertura y diseño experimental adecuado, los ensamblajes con referencia tampoco son capaces) y las regiones complejas suelen fragmentarse en más de una secuencia, necesitando de procesos alternativos y complementarios para eliminar fragmentos quiméricos y/o redundantes.

**Contig:** Secuencia de nucleótidos de longitud mayor al tamaño de las lecturas obtenidas, resultando del ensamblaje de más de una secuencia[87].

**Contig N50:** Se define como una media estadística propuesta para evaluar el rendimiento de un ensamblaje (genómico principalmente). El contig N50 es entonces la longitud media sobre la cual se contiene el 50% de la longitud total ensamblada, en contigs de igual o mayor longitud a este valor[90]. Dicho de otra manera, si ordenamos la longitud de los contigs en orden descendente, el contig N50 se corresponde a la longitud que, sumada a las anteriores contiene el 50% de la longitud total ensamblada. Es considerada, por lo tanto, como una forma indirecta de medir el grado de fragmentación de un ensamblaje.

**Cobertura de kmer:** Los ensambladores más comunes por algoritmos de Bruijn Graph (como el empleado en este trabajo) fragmentan las lecturas en trozos más cortos conocidos como kmeros (subsecuencias de longitud k). Estos kmeros constituyen las secuencias “semilla” sobre las que se ensamblan los fragmentos más largos[89,91]. Entonces, la cobertura de kmero se refiere al número de kmeros promedio usados en el ensamblaje de un contig individual.

**Cobertura de secuencia o RDC (*Read Depth Coverage*):** Se refiere al número de lecturas completas con similitud igual o superior a la establecida al contig o secuencia de referencia[89]. Dicho de otra forma, se corresponde al número de lecturas obtenidas de una región, transcrito, gen o contig específico.

**Scaffold:** Unión de dos o más contigs por estimaciones *in silico* de la distancia teórica entre ellos[87]. Para este proceso se utilizan normalmente lecturas pareadas de fragmento no solapante (conociendo así la distancia aproximada entre fragmentos), secuencias de referencia largas (incluidos genomas) y lecturas largas de baja cobertura.

**Complejidad lingüística:** Como una novela redactada por un escritor experimentado y un estudiante aprendiz, el texto del escritor experimentado contendrá un léxico más abundante y variado (menos repetitivo, mayor complejidad lingüística) mientras que el texto del estudiante será más repetitivo y de un léxico más limitado (más repetitivo, menor complejidad lingüística). En genética, la complejidad lingüística es directamente proporcional a la composición nucleotídica,

zonas altamente repetitivas son consideradas como de menor complejidad lingüística y viceversa, el %G+C es una variable comúnmente usada como indicativo de esta complejidad[92].

**Enriquecimiento funcional por ontología de genes o enriquecimiento de GO, (*Gene Ontology enrichment*):** Con el paso de los años y el avance en el conocimiento de la biología de los microorganismos, ha sido posible validar la influencia o grado de implicación de gran cantidad de genes en procesos biológicos o rutas metabólicas específicas. De esta manera, se asigna un número de identificación (GO ID) a cada gen, correspondiéndose con él, o los procesos en los que se tiene evidencia de su participación. Existen actualmente diversas herramientas que permiten el cálculo del enriquecimiento funcional general de un organismo basados en este principio[93,94].

Referentes a los análisis de transcriptómica, las medidas más populares de cuantificación de transcritos por RNAseq son el RPKM (*Reads Per Kilobase per Million*), el FPKM (*Fragments Per Kilobase per Million*) y TPM (*Transcripts Per Kilobase per Million*) y su objetivo principal es la representación de la abundancia relativa de cada transcrito en proporciones numéricas comparables entre distintas técnicas, replicas o situaciones biológicas[95,96]. Para ello, se consideran dos pasos (principalmente) de normalización, primero por la profundidad de la secuenciación inherente a la técnica e independiente de la situación biológica y segundo por la longitud de los transcritos en el sentido de que, por probabilidad simple, transcritos más largos contendrán más secuencias que los más cortos, independientemente de su nivel de expresión. El RPKM y el FPKM se calculan de la misma forma, con la única diferencia que este último es específico de RNAseq con lecturas pareadas, donde ambas lecturas corresponden a un solo fragmento, y por lo tanto se cuentan solo una vez[97]. Además, diferencia del RPKM y el FPKM, para calcular el TPM primero se debe normalizar por la longitud de los transcritos (en kb) y después por la profundidad de secuenciación.

**Efecto *batch*.** El efecto *batch* se define como una de las principales fuentes de variabilidad en experimentos de NGS, tales como el RNAseq. Este efecto puede ser originado por: condiciones de laboratorio, diferente origen de las muestras, diferencias de manipulación (entre personal) y diferentes técnicas experimentales. La detección y corrección de este efecto es indispensable para el correcto análisis de datos, y los algoritmos actualmente disponibles para este fin, se basan en el modelado y correlación de las posibles fuentes de errores intra e inter grupos[98].



## 2 OBJETIVOS

---

## 2.1 Objetivo general

Analizar mediante herramientas bioinformáticas la variabilidad genómica y genética de *Trypanosoma cruzi* durante el proceso de infección y el desarrollo de la enfermedad de Chagas, para dilucidar y contribuir al entendimiento de la respuesta inmune y las rutas metabólicas implicadas.

## 2.2 Objetivos específicos

1. Secuenciar y ensamblar los genomas de las cepas Bug2148 e Y, mediante técnicas de secuenciación masiva de segunda y tercera generación, para determinar las diferencias en contenido genético potencialmente relacionadas a las diferencias de patogenicidad en modelos *in vivo* e *in vitro*.
2. Determinar el transcriptoma completo de la cepa Sylvio X10, cuyo genoma estaba disponible en bases de datos, en el estado tripomastigote mediante RNAseq, para determinar el perfil transcriptómico de la forma infectiva, así como los principales mecanismos de maduración del ARNm.
3. Analizar mediante miARNseq el perfil de expresión de miARNs durante el proceso inicial de infección por *T. cruzi* en macrófagos peritoneales de ratón BALB/c, para determinar los potenciales mecanismos de regulación de la respuesta inmune por inhibición de la expresión génica.
4. Identificar el perfil de expresión génica del huésped en la fase crónica de la enfermedad de Chagas inducida por diferentes cepas de *T. cruzi*, para determinar las rutas biológicas potencialmente involucradas en la CCC.

### 3 MATERIALES Y MÉTODOS

---

### 3.1 Cultivo del parásito

Las cepas de *T. cruzi* Sylvio X10, CI Brener y VFRA usadas en este trabajo fueron facilitadas por el Dr. M. Miles (London School of Hygiene and Tropical Medicine, London, UK) a través de la red europea ChagasEpiNet (Séptimo Programa Marco de la U.E.). La cepa Y fue obtenida de la Harvard Medical School, Boston, Massachusetts, USA. Las cepas CI Brener y VFRA fueron inicialmente aisladas del insecto vector (*Triatoma infestans*) mientras que Sylvio X10 e Y fueron aisladas de pacientes infectados.

Los parásitos en el estado tripomastigote de las cepas Y, Bug2148 cl1 y Sylvio X10 usados para la secuenciación de ADNg, infección *in vitro* y RNAseq (transcriptoma de Sylvio X10), fueron obtenidos de la forma que se describe a continuación.

Se cultivaron células Vero (de riñón de Mono verde africano) a 37 °C y ambiente de CO<sub>2</sub> al 5% en medio RPMI suplementado con 5% de suero fetal bovino (SFB), 100 UI/ml de mezcla de antibióticos, 10 µg/ml de estreptomina y 2 mM de glutamina. Después de 4 días (y más del 80% de confluencia) la mono-capa celular fue infectada con una relación parásito-célula 5:1. Después de 4 días se recuperó el sobrenadante y se lavó con PBS estéril, los restos celulares y amastigotes fueron removidos por centrifugación a 1000 g por 5 minutos y los tripomastigotes fueron recolectados por centrifugación a 1600 g durante 10 minutos.

Los parásitos en el estado tripomastigote metacíclico sanguíneo de las cepas Y, CI Brener, VFRA y SylvioX10 usados en este trabajo para la infección *in vivo* fueron obtenidos como se describe a continuación.

La cepa Y fue mantenida por inyección semanal intraperitoneal en ratones BALB/c en las instalaciones de bioseguridad de nivel 3 del CBMSO. Las cepas VFRA, CI Brener y Sylvio X10 fueron mantenidos en ratones BALB/c tras inyección intraperitoneal y la inmunosupresión del huésped mamífero siguiendo el protocolo descrito por Calabrese y col. en 1996[99], 5 días después de la infección los ratones son inyectados con una dosis de ciclofosfamida (200 mg/kg de peso).

### 3.2 Ratones y declaración ética

Ratones hembra de 6-8 semanas de edad de la cepa BALB/c fueron adquiridos de la empresa “Charles River Laboratories, España” y mantenidos en condiciones libres de patógenos en el servicio de animalario del Centro de Biología Molecular Severo Ochoa (CBMSO).

Todos los experimentos fueron realizados bajo un estricto seguimiento de la legislación del parlamento europeo y el consejo de la unión europea relativa a la protección de los animales utilizados para fines científicos (2010/63/EU). El protocolo de tratamiento de los animales fue aprobado por el comité de ética de legislación de la comunidad de Madrid (permiso PROEX 148/15). Los animales tuvieron en todo momento acceso ilimitado a comida y agua. Fueron sacrificados en una cámara de CO<sub>2</sub> procurando siempre la disminución de su sufrimiento.

### **3.3 Obtención de macrófagos intraperitoneales e infección *in vitro* e *in vivo***

Se inyectó intraperitonealmente 1 ml de tioglicolato al 10% en PBS (Gibco, Grand Island, NY) a cada ratón. Después de 4 días, los ratones fueron sacrificados en una cámara de CO<sub>2</sub> y los macrófagos fueron colectados mediante lavado con 10 ml de PBS frío y centrifugación a 260g durante 10 min a 4 °C. Posteriormente se cultivaron *overnight* 10 millones de células en placas p100 con medio RPMI (descrito anteriormente) a 37 °C y ambiente de CO<sub>2</sub> al 5%. Al día siguiente, las células no adheridas y restos celulares fueron eliminadas mediante 3 lavados con PBS.

Infección *in vitro*. Los macrófagos cultivados en placa fueron infectados (en el laboratorio de bioseguridad nivel 3 del CBMSO) con una relación parásito célula 5:1. Después de 1, 6 y 24 horas post-infección, las placas fueron lavadas 3 veces con PBS (incluidos los controles no infectados de cada tiempo) y reservadas a -70 °C para la extracción de ARN total.

La infección *in vivo* se realizó con la inoculación de 50 parásitos en estado tripomastigote metacíclico sanguíneos de las cepas Y, CI Brener, Sylvio X10 y VFRA en ratones BALB/c hembra de 6-8 semanas de edad y 6 réplicas. Los ratones fueron sacrificados 6 meses post-infección (en la etapa crónica de la enfermedad) en cámara de CO<sub>2</sub>. Posteriormente, el corazón fue extraído, lavado 3 veces con PBS y resguardado a -70 °C para la extracción de ARN total.

### **3.4 Extracción de ARN total**

#### **3.4.1 ARN total de la infección *in vitro* a 1, 6 y 24 horas**

El RNA total (incluidos miARNs) de las infecciones *in vitro* (incluidos controles no infectados) fueron extraídos con el kit “*RNeasy Mini Kit*” de Qiagen® lisando directamente sobre la placa con 700 µl de buffer de lisis y posteriormente tratando

con DNAasa I libre de RNAasas (Roche®). En todos los casos se evaluó la concentración y pureza del ARN extraído mediante absorbancia a 260 nm usando Nanodrop ND-1000 (Thermo Scientific). En todos los casos la relación A260/A280 se mantuvo en torno a 2. Además, se evaluó la integridad del material extraído mediante bioanalyzer Agilent 2100 y de la misma forma, en todos los casos el número de integridad de RNA o RIN (*RNA integrity number*) fue superior a 8.

#### **3.4.2 ARN total de la infección *in vivo* a 6 meses post-infección**

El ARN total de corazón fue aislado con el kit “*RNeasy Mini Kit*” de Qiagen® previa homogenización del tejido con ayuda del Polytron homogenizador PT 1200E. Se evaluó la concentración y pureza del ARN extraído mediante absorbancia a 260 nm usando Nanodrop ND-1000 (Thermo Scientific). La relación A260/A280 se mantuvo en torno a 2 y un RIN superior a 8.

#### **3.4.3 ARN total de tripomastigotes de la cepa Sylvio X10**

El RNA total de la cepa Sylvio X10 en la forma tripomastigote fue extraído de 10 millones de parásitos usando el kit “*RNeasy Mini Kit*” de Qiagen® y posteriormente tratado con DNAasa I libre de RNAsas (Roche®). Se evaluó la concentración y pureza del ARN extraído mediante absorbancia a 260 nm usando Nanodrop ND-1000 (Thermo Scientific). La relación A260/A280 se mantuvo en torno a 2 y un RIN superior a 8.

### **3.5 Extracción de ADN**

Las diferentes tecnologías de secuenciación de ADN empleadas en este trabajo (PacBio e illumina) para el ensamblaje de los genomas de las cepas Bug2148 e Y, respectivamente, requieren condiciones de integridad del ADN en diferente grado. En el caso de PacBio, es necesario que el ADN no se encuentre fragmentado en trozos menores a 30-40 Kb, mientras que para Illumina fragmentos inferiores a 10 Kb suelen ser suficientes (los fragmentos secuenciados usualmente no superan 1 kb).

Por lo tanto, el ADNg de 10 millones de tripomastigotes de la cepa Y fue extraído usando el kit “*High Pure PCR Template Preparation Kit*” de Roche®. El ADNg de la cepa Bug2148 (20 millones de parásitos) fue extraído con el clásico método fenol-cloroformo omitiendo el último paso de precipitación del material genético por

centrifugación, en su lugar, el pellet de ADN fue extraído. Ambas muestras fueron tratadas con RNasa I libre de DNAsas (Roche®) y cuantificadas por absorbancia a 260 nm con Nanodrop ND-1000 (*Thermo Scientific*), mostrando en ambos casos una relación A260/A280 en torno a 2.

### **3.6 Secuenciación, ensamblaje y análisis de los nuevos genomas de las cepas Bug2148 e Y**

El genoma de Bug2148 fue secuenciado con la tecnología PacBio en el “Norwegian Sequencing Centre”, en Oslo, Noruega. La cepa Y fue secuenciada (fragmentos de 900 pb) con la tecnología Illumina (MiSeq) por la Unidad de Genómica en el Parque Científico de Madrid, España. En ambos casos, los servicios de secuenciación siguieron los protocolos establecidos por ambas tecnologías para la preparación de las librerías y posterior secuenciación. Con base a estimaciones previas del tamaño total del genoma, en torno a 115 y 106 Mb para la cepa Y y Bug2148[26,100], respectivamente, se estimó una profundidad de secuenciación de 100X.

Se generaron lecturas pareadas no solapantes en formato 2x300 pb y de 8-15 kb de longitud por las diferentes tecnologías para Y y Bug2148, respectivamente.

Las lecturas cortas de Illumina fueron filtradas y recortadas por calidad con el software FASTQC[101] y Prinseq-lite[102] con los siguientes parámetros, longitud mínima 50 nucleótidos, calidad media mínima de 25 y longitud máxima 250 nucleótidos.

Las lecturas generadas por PacBio fueron filtradas por calidad mínima de 0.8 y longitud mínima de 500 bases.

Ensamblaje. Las lecturas filtradas de la cepa Y fueron ensambladas con el software SPAdes (v3.9.0)[91] usando la longitud de kmeros 21, 33, 55, 77, 99 y 127. Mientras que el genoma de Bug2148 fue ensamblado con el software HGAP v3 (*SMRT Analysis Software v2.3.0*)[103] estableciendo la longitud de la secuencia semilla en 6 kb y 15X de cobertura mínima.

La estadística del ensamblaje fue calculada con el software Biopieces tool kit[104]. El porcentaje de G+C (%GC) fue calculado para cada contig ensamblado de manera progresiva (ventana= 1 kb, paso= 500 bases) mediante scripts propios y de uso público[105].

La redundancia de secuencia ensamblada fue eliminada mediante la eliminación de contigs con cobertura menor a 10 (cobertura de kmero) e intra-similitud mayor o

igual al 90% detectada con el software CD-HIT[106] y scripts propios en lenguaje Python.

### **3.6.1 Genes de las Trans-sialidasas, secuencias teloméricas y secuencias repetitivas en los ensamblajes**

Las familias de TSs fueron identificadas mediante los resultados de BLAST y *Markov Cluster Algorithm* (MCL), las repeticiones teloméricas y regiones complejas ( $0.35 \leq \%G+C \leq 0.65$ ) fueron determinadas mediante la función `patscan_seq` (sin admitir *mismatches*) de Biopieces, IGV y scripts propios de Python.

### **3.7 Predicción genética y anotación genómica**

La predicción genética fue realizada con el algoritmo Prodigal (v2.6.3)[107] en su versión de uso local, ajustando la predicción exclusiva de ORFs completas con la tabla de traducción eucariota estándar. Las familias genéticas fueron determinadas con MCL y sus parámetros por defecto[108]. La función de cada una de las familias genéticas encontradas fue establecida por el mejor hit de BLAST (*Basic Local Alignment Search Tool*: e-value menor o igual a  $1e-5$  y porcentaje de identidad mayor al 50%) contra la base de datos no redundante (*nr*) del NCBI (*National Center for Biotechnology Information*). Las anotaciones genómicas fueron inspeccionadas manualmente en medida de lo posible con el software *Integrative Genomics Viewer* (IGV)[109]. Los genes de copia única fueron identificados de los resultados de MCL.

### **3.8 Secuenciación y ensamblaje del transcriptoma de Sylvio X10 en su forma tripomastigote**

La secuenciación se realizó por la Unidad de Genómica del Parque Científico de Madrid, España con la tecnología Illumina (Miseq). Las librerías de secuenciación fueron preparadas siguiendo los protocolos establecidos por la tecnología sin modificaciones. El ARNm fue seleccionado por la complementariedad nucleotídica de sus cadenas poli-A con oligonucleótidos de timina en cromatografía por columna y fragmentado a una longitud promedio de 450-500 bases. Se sintetizó el ADNc de este ARNm y se prepararon las librerías de secuenciación.

Se generaron lecturas pareadas de 100 nucleótidos de longitud y fueron filtradas por calidad mínima promedio (25) y longitud mínima de 50 bases con el software FASTQC[101] y Prinseq-lite[102].



Las lecturas filtradas fueron mapeadas a la referencia con el software Bowtie2[110], el ensamblaje del transcriptoma y la cuantificación de los transcritos fue realizada con el software StringTie[111]. La corrección de la secuencia genómica fue realizada con Pilon[112].

La identificación de la función más probable de las proteínas codificadas en el transcriptoma se determinó con el programa Blast2GO[113], BLAST contra la base de datos *nr* del NCBI y scripts propios en lenguaje Python.

Los sitios de trans-splicing y poli-adenilación fueron identificados como se describe en el trabajo publicado por Rastrojo y col.[39] y la versión de uso local de WebLogo[114].

Adicionalmente, las herramientas bioinformáticas samtools[115] e IGV[109] fueron empleadas para manipular, visualizar y analizar los datos de secuenciación.

### **3.9 Disponibilidad de datos de secuenciación**

Los datos de secuenciación de los genomas ensamblados de Bug2148 e Y, fueron depositados en las bases de datos del NCBI con los números de acceso NMZN000000000 y NMZO000000000, respectivamente.

Los datos de secuenciación del transcriptoma de Sylvio X10 fueron depositados en las bases de datos del NCBI con el número de acceso PRJNA546488, así como en el repositorio del Trytripdb.

### **3.10 MiARNseq a 1,6 y 24 horas post-infección, secuenciación y análisis**

La fracción de RNA total correspondiente a miARNs (15-30 bases) fue purificado mediante exclusión por tamaño en geles de poliacrilamida, posteriormente se prepararon las librerías de secuenciación siguiendo los protocolos específicos de la tecnología Illumina (TruSeq) sin modificaciones, por la Unidad de Genómica del Parque Científico de Madrid, España y se generaron lecturas de 14-50 bases de longitud.

Las lecturas de secuenciación fueron filtradas por calidad promedio mínima superior a 25 y longitud mínima de 10 bases con el software FASTQC[101] y Prinseq-lite[102]. Posteriormente estas lecturas filtradas fueron alineadas al genoma de ratón (mm9) descargada de la base de datos de la *University of California Santa Cruz* (UCSC) con Bowtie[116], contadas con el programa HTSeq-count[117] y la última versión del archivo de anotación de miARNs (11 de abril de 2019) en el genoma de ratón, descargado de la base de datos de mirbase.org.

### **3.11 Expresión diferencial de miARNs en macrófagos en respuesta a la infección por *T. cruzi* e identificación de los potenciales ARNm diana**

El análisis de expresión diferencial entre los diferentes tiempos de análisis y su control no infectado se realizó con el programa DESeq2[118] y scripts propios en lenguaje Python.

Las dianas de los miARNs con expresión diferencial significativa, así como su análisis de enriquecimiento funcional (1 y 6 horas post-infección “h.p.i”) fueron determinados mediante el software Diana-mirPath v3[119].

Las interacciones entre las proteínas codificadas por las dianas de los miARNs con expresión diferencial a 24 h.p.i, fueron determinadas mediante la plataforma “*ELIXIR Core Data Resources*”[120] y la versión 11 de la aplicación String (<https://string-db.org/>).

### **3.12 Identificación de potenciales miARNs de *T. cruzi* en macrófagos intraperitoneales de ratón BALB/c infectados con la cepa Y**

Las lecturas de miARNseq no alineadas al genoma de ratón fueron aisladas y alineadas con Bowtie2[110] al genoma de *T. cruzi* cepa Y ensamblado en este trabajo. En consecuencia, las lecturas alineadas fueron extraídas y consideradas como material genético *bona fide* del parásito. Los *clusters* con longitud teórica de miARNs y RDC (cobertura de secuencia) mayor a 1000 fue identificada mediante scripts propios en lenguaje Python y R.

La predicción de precursores y miARNs maduros de *T. cruzi* se realizó mediante el algoritmo MirDeep2[121] en las 3 etapas de infección (1, 6 y 24 h.p.i). La validación de la presencia de estos miARNs en vesículas extracelulares del parásito (vesículas de epimastigotes, amastigotes y célula completa de la cepa Dm28c) se realizó usando los datos de acceso público de Bayer-Santos y col.[122] depositados con el número de acceso GSE54111.

### **3.13 ARNseq de macrófagos intraperitoneales de ratón BALB/c infectados con *T. cruzi*, 24 horas post-infección**

La secuenciación fue realizada en el “*Centre de Regulació Genòmica*” (CRG) de Barcelona, España con la tecnología Illumina (Miseq). Las librerías de secuenciación fueron preparadas siguiendo los protocolos establecidos por la tecnología sin modificaciones. El ARNm fue seleccionado por la complementariedad nucleotídica de sus cadenas poli-A con oligonucleótidos de timina en cromatografía por columna y fragmentado a una longitud promedio de 450-500 bases. Se sintetizó el ADNc de este ARNm y se prepararon las librerías de secuenciación.

Se generaron lecturas pareadas de 75 nucleótidos de longitud que fueron filtradas por calidad mínima promedio (25) y longitud mínima de 50 bases con el software FASTQC[101] y Prinseq-lite[102]. La contaminación de secuencia con adaptadores fue removida con el software Trimmomatic[123].

Las lecturas filtradas fueron alineadas y analizadas con el genoma de ratón (mm9) de la base de datos de la (UCSC) siguiendo el nuevo protocolo “*Tuxedo*”[124] con algunas modificaciones, usando Hisat2 (en lugar de la versión 1), StringTie y DESeq2 (en lugar de Ballgown).

El análisis de enriquecimiento funcional se realizó con la versión online más reciente de “*WEB-based GENE SeT Analysis Toolkit*” (WebGestalt, disponible en marzo del 2019)[93] y la matriz de genes con expresión significativa ( $p$ -valor ajustado  $\leq 0.001$ ) usando el segundo logaritmo del *Fold of Change* ( $\log_2\text{Fold}$ ) respecto a la muestra control como sistema jerárquico. Las rutas metabólicas fueron analizadas con las herramientas disponibles de WikiPathways[125] y las bases de datos de la “*Kyoto Encyclopedia of Genes and Genomes*” (KEGG) de “*Kanehisa Laboratories*”[126].

### **3.14 ARNseq de corazón de ratones BALB/c infectados con *T. cruzi*, 6 meses post-infección**

La secuenciación fue realizada en el CRG de Barcelona, España con la tecnología Illumina (Miseq). Las librerías de secuenciación fueron preparadas siguiendo los protocolos establecidos por la tecnología sin modificaciones. El ARNm fue seleccionado por la complementariedad nucleotídica de sus cadenas poli-A con oligonucleótidos de timina en cromatografía por columna y fragmentado a una longitud promedio de 450-500 pb. Se sintetizó el ADNc de este ARNm y se prepararon las librerías de secuenciación.

Se generaron lecturas *single-end* de 75 nucleótidos de longitud y fueron filtradas por calidad mínima promedio (25) y longitud mínima de 50 bases con el software

FASTQC[101] y Prinseq-lite[102]. La contaminación de secuencia de adaptadores fue removida con el software Trimmomatic[123].

Las lecturas filtradas fueron alineadas y analizadas con el genoma de ratón (mm9) de la base de datos de la (UCSC) siguiendo el nuevo protocolo “*Tuxedo*”[124] con algunas modificaciones, usando Hisat2 (en lugar de la versión 1), StringTie y DESeq2 (en lugar de Ballgown).

El análisis de enriquecimiento funcional se realizó con la versión online más reciente de “*WEB-based GENE SeT AnaLysis Toolkit*” (WebGestalt, disponible en marzo del 2019)[93] y la matriz de genes con expresión significativa ( $p$ -valor ajustado  $\leq 0.001$ ) usando el segundo logaritmo del *Fold of Change* ( $\log_2\text{Fold}$ ) respecto a la muestra control como sistema de jerarquía. Los genes sin expresión significativa, pero *baseMean* mayor o igual a 100 fueron usados como background con expresión igual a 0. Las rutas metabólicas fueron analizadas con las herramientas disponibles online de WikiPathways[125] y las bases de datos de la “*Kyoto Encyclopedia of Genes and Genomes*” (KEGG) de “*Kanehisa Laboratories*”[126].

### 3.15 Estadística y figuras

Todos los análisis estadísticos fueron realizados con el software de uso libre R y Rstudio (V. 3.4.2). Las figuras fueron realizadas con Rstudio, Cytoscape (V.3.5.1) y formateadas con Illustrator CC (V.2017).

#### 4 RESULTADOS Y DISCUSIÓN

---

## 4.1 Secuenciación de dos nuevos genomas de *T. cruzi*

### 4.1.1 Genómica de *T. cruzi*, secuenciación y ensamblaje

Las cepas de *T. cruzi* más abundantes en el ciclo doméstico y por tanto más estudiadas en el proceso del desarrollo de la enfermedad en humanos y modelos *in vitro* en ratón pertenecen a las DTUs I,II,V y VI[127,128]. Sin embargo, en el momento de la publicación de los resultados obtenidos en este trabajo solamente se había secuenciado un genoma de la DTU II (Esmeraldo cl3) y ninguno de la DTU V. Por lo tanto, se seleccionaron las cepas Y (DTU II) y Bug2148 cl1 (DTU V) para ser secuenciadas, analizadas y comparadas con los genomas disponibles en bases de datos. Bug2148 fue seleccionada además por ser una cepa asociada a infección por transmisión vertical (de madres a hijos)[129], mientras que la cepa Y fue elegida por su importancia fenotípica en términos de virulencia, infectividad y desarrollo de la enfermedad en modelos *in vitro* e *in vivo*[130–132].

La cepa Y fue secuenciada con la tecnología Illumina, obteniendo aproximadamente 20 millones de lecturas pareadas de hasta 250 pb mientras que Bug2148 fue secuenciada con la tecnología Pacific Biosciences (PacBio) y se obtuvieron 757,037 lecturas con longitud media de 14 Kb. En ambos casos, las lecturas de secuenciación obtenidas se corresponden a aproximadamente 10 Gb de información y más de 100X de cobertura de los genomas haploides predichos[21,26].

Contigs de secuencia redundante (90% de similitud) y baja cobertura (menor a 10) fueron descartados de los ensamblajes de ambos genomas, obteniendo como resultado final 10,127 y 934 contigs para Y y Bug2148, respectivamente (**Tabla 2**).

Gracias a la principal ventaja que ofrece la tecnología PacBio, que permite la secuenciación de fragmentos genómicos largos (hasta 20 Kb) contrarrestando la naturaleza repetitiva y compleja de *T. cruzi*, el genoma de la cepa Bug2148 fue ensamblado en el menor número de contigs que cualquier otra cepa antes secuenciada, contiene el contig más largo secuenciado y ensamblado *de novo* para esta especie (1.3 Mb), el contig N50 más largo, el promedio de longitud más alto (a excepción de Cl Brener) y la tercera mejor cobertura. Además, teniendo en cuenta el número de bases ensambladas, el tamaño total predicho de su genoma y el número total de fragmentos obtenidos, esta secuencia es la referencia genómica más completa y menos redundante hasta el momento.

Como se esperaba respecto a cepas anteriormente ensambladas, el porcentaje total de G+C es ligeramente superior al 50%.

**Tabla 2.** Resumen de los genomas de *T. cruzi* actualmente secuenciados y ensamblados. Se subrayan los genomas obtenidos en este trabajo.

Cepa	DTU	Tamaño (Mb)	GC (%)	Contig N50 *	Longitud Máxima (pb)	Longitud Mínima (pb)	Longitud promedio (pb)	Contigs totales	Ensamblaje*	Cobertura (%)	Tecnología de secuenciación
Dm28c	I	27.34	53.33	78,389	462,134	2,004	22,601	1,210	Scaffold	63	454*
JrcI4	I	41.48	52.45	83,591	828,981	200	2,709	15,312	Scaffold	69	454
Sylvio X10	I	38.58	52.91	2,307	72,500	202	1,428	27,019	Contigs	30	454 + illumina
<u>Y</u>	<u>II</u>	<u>39.34</u>	<u>51.43</u>	<u>11.782</u>	<u>304.87</u>	<u>500</u>	<u>3.885</u>	<u>10.127</u>	<u>Contigs</u>	<u>71</u>	<u>Illumina</u>
Esmeraldo cl3	II	38.08	52.09	66,229	483,664	200	2,409	15,803	Scaffold	60	454
<u>Bug2148</u>	<u>V</u>	<u>55.22</u>	<u>51.63</u>	<u>196.760</u>	<u>1.305.792</u>	<u>585</u>	<u>59.129</u>	<u>934</u>	<u>Contigs</u>	<u>68</u>	<u>PacBio*</u>
Tula	VI	83.51	50.63	7,772	242,476	200	1,826	45,711	Scaffold	50	454
BNEL	VI	32.52	43.94	870,934	2,371,736	77,958	793,392	41	Chr*	14	454
BEL	VI	32.52	40.35	870,934	2,371,736	77,958	793,391	41	Chr*	14	454
B7	VI	38.65	52.87	25,781	335,615	235	2,302	16,783	Contigs	30	454 + illumina

\*Chr: Scaffolds largos simulando el tamaño aproximado de cromosomas.

\*454: Roche 454

La reconstrucción de cromosomas enteros de hasta varios Mb de longitud no es posible mediante lecturas cortas de 250 pb como las que genera la tecnología illumina, lo que ocasiona obtener genomas muy fragmentados incluso en el rango de miles de trozos. Por otra parte, la gran cantidad de lecturas que esta tecnología es capaz de generar, permitió obtener para la cepa Y el genoma de más alta cobertura obtenido hasta el momento (71X) y el tercer mejor en promedio de longitud (3.8 Kb) solo después de Bug2148 y Cl Brener.

Uno de los principales problemas en genomas ensamblados altamente fragmentados independientemente del nivel intrínseco de complejidad, es que no se puede determinar su contenido genético total con exactitud. En este sentido, Arner y col.[133] sugieren el uso de genes multicopia altamente conservados en *T. cruzi* independientemente de la DTU, como una medida indirecta del nivel de redundancia e información contenida de un ensamblaje de este tipo. Por ejemplo, se calcula que existen 50 copias de la proteína Monoglicérido lipasa a lo largo del genoma, sin embargo, solo se encuentra anotada una vez en el genoma de Sylvio X10 y Cl Brener, ninguna en el resto de cepas secuenciadas. Esta misma proteína ha sido encontrada y anotada 4 veces en el genoma de la cepa Y y 30 en el de la cepa Bug2148, confirmando el nivel de completitud de ambos genomas y las grandes diferencias de alcance entre ambas tecnologías, al menos en proyectos relacionados a *T. cruzi*.

#### 4.1.2 Predicción genética y análisis funcional

El próximo paso fue la determinación del contenido genético en los genomas ensamblados, cantidad y clasificación.

Además del filtro por redundancia a nivel de secuencia y cobertura de secuenciación, contigs menores a 500 nucleótidos de longitud y sin contener marcos de lectura abiertos u ORFs (*Open Reading Frames*) fueron removidos de ambos ensamblajes. ORFs menores a 50 aminoácidos no fueron considerados en análisis posteriores. En total, fue posible determinar 33.306 y 20.058 ORFs en Bug2148 e Y, respectivamente, este último en acuerdo con estimaciones previas para cepas consideradas como ancestrales. Sin embargo, el número de ORFs encontradas en Bug2148 es mucho mayor a lo estimado en *T. cruzi*, incluso para cepas híbridas.

El análisis funcional de los genes predichos se realizó en dos etapas, primero las familias genéticas fueron identificadas mediante self-Blastp y agrupación con el algoritmo MCL, posteriormente la función más probable de cada gen fue determinada mediante BLASTp contra la base de datos no redundante de protozoarios del NCBI. Genes sin función conocida (o hipotética conservada) fueron anotadas como “Proteínas hipotéticas”.

Después de este análisis funcional de clasificación y agrupación, fue posible determinar la existencia de 10.549 y 10.674 familias genéticas en Bug2148 e Y, respectivamente. Tal como se ha descrito anteriormente, aproximadamente el 50% del contenido genético en *T. cruzi* corresponde a funciones desconocidas (hipotéticas) **Tabla 3**.

**Tabla 3.** Porcentaje de genes en diferentes cepas de *T. cruzi* con función desconocida o hipotética.

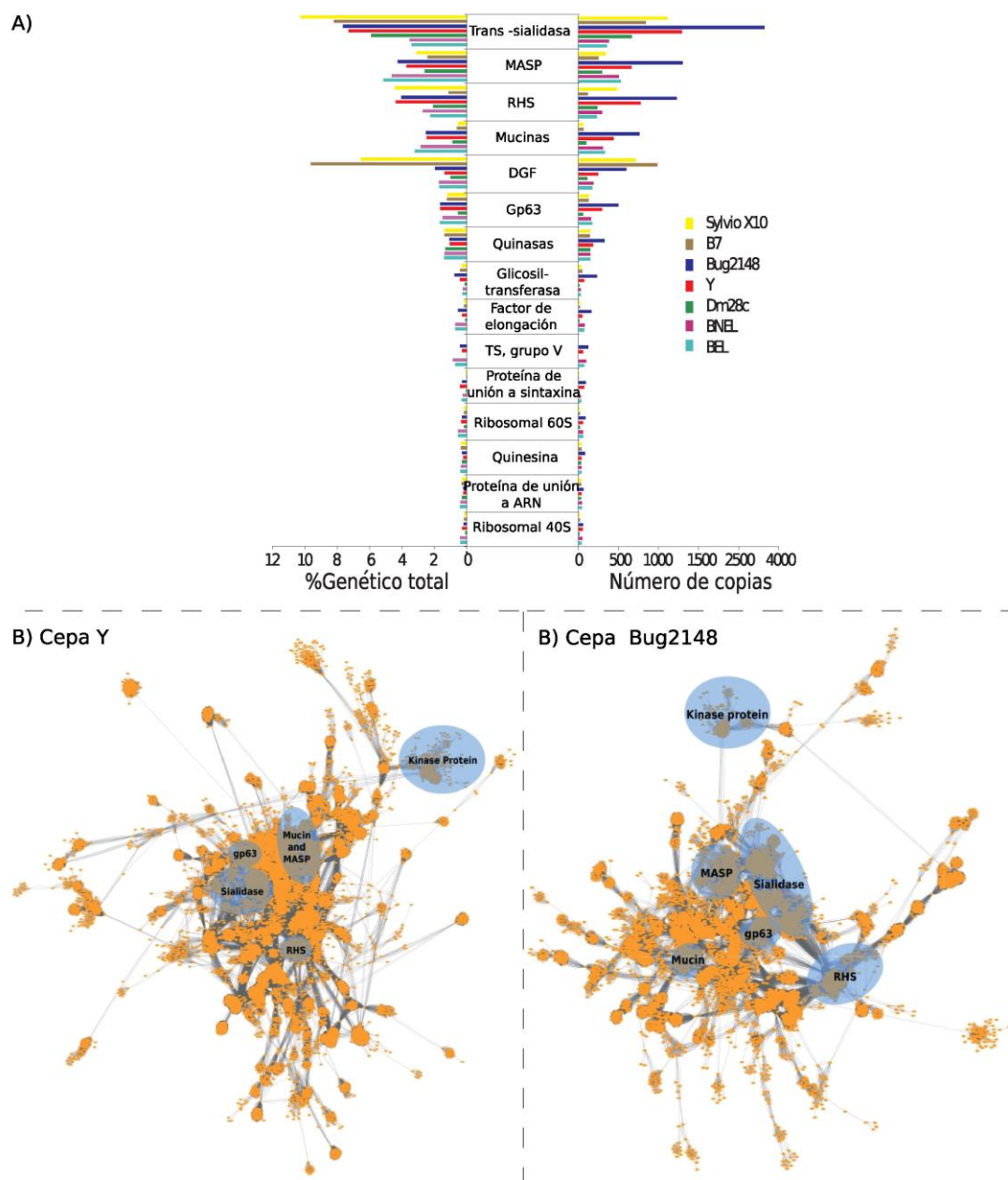
Cepa	Proteína (%) Hipotética
Dm28c	64,26
Sylvio X10	49,50
<u>Y</u>	<u><b>56,94</b></u>
<u>Bug2148</u>	<u><b>53,25</b></u>
BEL	51,53
BNEL	51,55
B7	50,60

Posteriormente, se realizó un comparativo de las familias genéticas identificadas de los nuevos genomas ensamblados con algunas otras cepas disponibles en las bases



de datos, las familias genéticas en estas últimas fueron determinadas a partir de sus archivos de anotación (genomas no anotados fueron excluidos del análisis).

En la **Figura 7A** se muestran las 15 familias genéticas más abundantes en *T. cruzi*, las cuales además constituyen aproximadamente el 25% del contenido genético total (24,36% en Y y 26,06% para Bug2148).



**Figura 7.** Familias genéticas en *T. cruzi*. **A)** En el panel superior se muestran las 15 familias más abundantes encontradas, a la derecha el número de copias y a la izquierda el porcentaje de cada familia respecto al total. **B)** Agrupación del contenido genético total en la cepa Y, **C)** Agrupación del contenido genético total en la cepa Bug2148. Se resaltan en círculos azules los grupos conocidos más representativos.

En resumen, se obtuvieron 33.306 y 20.058 ORFs en los nuevos genomas, de los cuales fue posible identificar 10.549 y 10.674 familias (Bug2148 e Y, respectivamente). Como era de esperar en esta especie, poco más del 50% del contenido genético total tiene función desconocida o hipotética. Del 50% restante, la mitad se compone solo de 15 familias (**Figura 7A**), aunque es posible identificar un total de 1.589 funciones conocidas entre las cepas incluidas en este análisis.

Más del 80% de los genes de función hipotética se corresponden con genes altamente conservados entre distintas cepas. Fue posible identificar más de 8.500 grupos en esta condición de hasta 700 miembros, por lo que podrían constituir un grupo genético de análisis con gran relevancia.

Aunque en la cepa Bug2148 se encontraron más de 13 mil genes adicionales que en la cepa Y, el total de familias fue prácticamente el mismo (solo 125 más en Y), lo que en general confirma la teórica naturaleza híbrida de Bug2148 con un repertorio genético más amplio, al menos duplicado. Además, es posible determinar cierto grado de divergencia genética entre ambas cepas respecto a las 5 secuencias proteicas más abundantes (**Figura 7B,C**), que además han sido descritas como uno de los principales factores de virulencia e involucradas en el desarrollo de la enfermedad.

En concordancia con publicaciones anteriores que describen la composición genética de *T. cruzi*, los genes más abundantes se corresponden con aquellos que codifican principalmente para proteínas de superficie tales como Trans-sialidasas (TS), Mucinas, proteínas de superficie asociadas a mucinas o MASP por sus siglas en inglés (*Mucin associated surface protein*), otras asociadas a mecanismos de expresión como RHS (*Retrotransposon Hot Spot*) y proteínas cuya función específica continua siendo desconocida[25,30,134].

La comparación del contenido genético de distintas cepas de *T. cruzi* secuenciadas y ensambladas actualmente, tanto por número de copias como en porcentaje respecto al contenido total, reveló similitudes y diferencias interesantes. Por ejemplo, el número estimado de genes codificantes de actividad Trans-sialidasa en cepas ancestrales (DTUs I y II) es de 1.400[135], coincidentemente se obtuvieron un total de 1.419 genes con esta actividad en la cepa Y, mientras que en Sylvio X10 se han anotado 1.321. En contraste, en la cepa Bug2148 fue posible determinar casi el doble, 2.325 (**Figura 7A**), sin embargo, el porcentaje respecto al total que esta extensa familia representa, es variable entre las distintas cepas. Aproximadamente el 8% de las proteínas codificadas en Bug2148, Y y B7 tienen actividad trans-sialidasa, mientras que en Sylvio X10 constituyen más del 10% y 6% en Dm28c y los

2 haplotipos de CI Brener (*Brener Esmeraldo-like*: BEL y *Brener non Esmeraldo-like*: BNEL).

Por otra parte, la abundancia de los genes codificantes de MASP relacionados con procesos de evasión de la respuesta inmunitaria del huésped y que se estima constituyen 6% del contenido genético total[51], es más estable entre las cepas de estudio, situándose alrededor del 3,5%, Dm28c: 3,5%, CI Brener: 3,5%, B7: 3,0%, Y: 3,7% y Bug2148: 3,9%.

Sin embargo, para otra de las familias multi-génicas más abundantes en esta especie la situación es un poco diferente. Genes codificantes de RHS, una proteína asociada a regiones teloméricas y procesos de replicación del DNA son tan abundantes como los genes MASP en Y, Bug2148, Sylvio X10, pero mucho menos abundantes en la cepa avirulenta *T. cruzi marinkellei* B7.

Este análisis demostró además que la familia multi-génica de mucinas es la cuarta más abundante, no la segunda como se había supuesto anteriormente basados exclusivamente en el contenido genético de la cepa más común de referencia, el híbrido CI Brener. Además, existen grandes diferencias de abundancia entre las diferentes cepas. Así una de las cepas más virulentas descritas en *T. cruzi* (Y) con los híbridos Bug2148 y CI Brener son las cepas con mayor contenido de esta proteína (hasta 750 copias en Bug2148), mientras que en las cepas Sylvio X10, B7 y Dm28c el contenido es dramáticamente menor (hasta un séptimo).

De acuerdo con estimaciones previas, la quinta familia genética más abundante en *T. cruzi* es DGF (*Dispersed Gene Family*)[51,136], una proteína de superficie diferencialmente expresada a lo largo del ciclo de vida del parásito. Sin embargo, se encontró que la cantidad estimada (565) es altamente variable entre cepas, cumpliéndose solo en algunas de las cepas (Sylvio X10, Bug2148 y B7) donde constituye hasta el 10% del total (B7), mientras que para el resto la cantidad total no llega a la mitad (Y, Dm28c y CI Brener). Además, debido a la variabilidad a nivel de secuencia de este tipo de proteína (85-95%) generada por la susceptibilidad a recombinación genética constante[136,137], se había propuesto una subclasificación formada por al menos 3 grupos principales, sin embargo a partir de la nueva información disponible se determinó que la complejidad de esta familia es superior, especialmente en Bug2148.

Otra de las familias multigénicas importantes en *T. cruzi* y que forman parte del grupo de las 15 más abundantes, son las proteínas de unión a syntaxina (*Syntaxin-binding proteins*). Forman parte del grupo de las SNAREs por sus siglas en inglés (*Soluble N-ethylmaleimidesensitive factor attachment protein receptors*) y están involucradas en distintos procesos de tráfico intra-celular y la evasión del sistema

inmune en el huésped. De acuerdo con los resultados obtenidos, esta proteína es más abundante en las cepas híbridas y en la cepa Y.

En general, los resultados de este análisis comparativo confirman que, a día de hoy, el contenido y diversidad de las 6 familias genéticas más abundantes en *T. cruzi* (TS, MASP, RHS, Mucins, DGF y gp63) es específico de cada cepa, dependiente del grado de completitud de su genoma, de la complejidad a nivel de secuencia y de la plasticidad genómica inherente a la especie y por lo tanto independientemente de su clasificación taxonómica. Debido al importante papel que desempeñan estas familias genéticas en múltiples procesos asociados al desarrollo de la enfermedad y durante el ciclo de vida del parásito, es importante seguir analizando la relación entre su diversidad entre cepas y el efecto biológico inherente, pero sin olvidar su complejidad y grado de evolución constante.

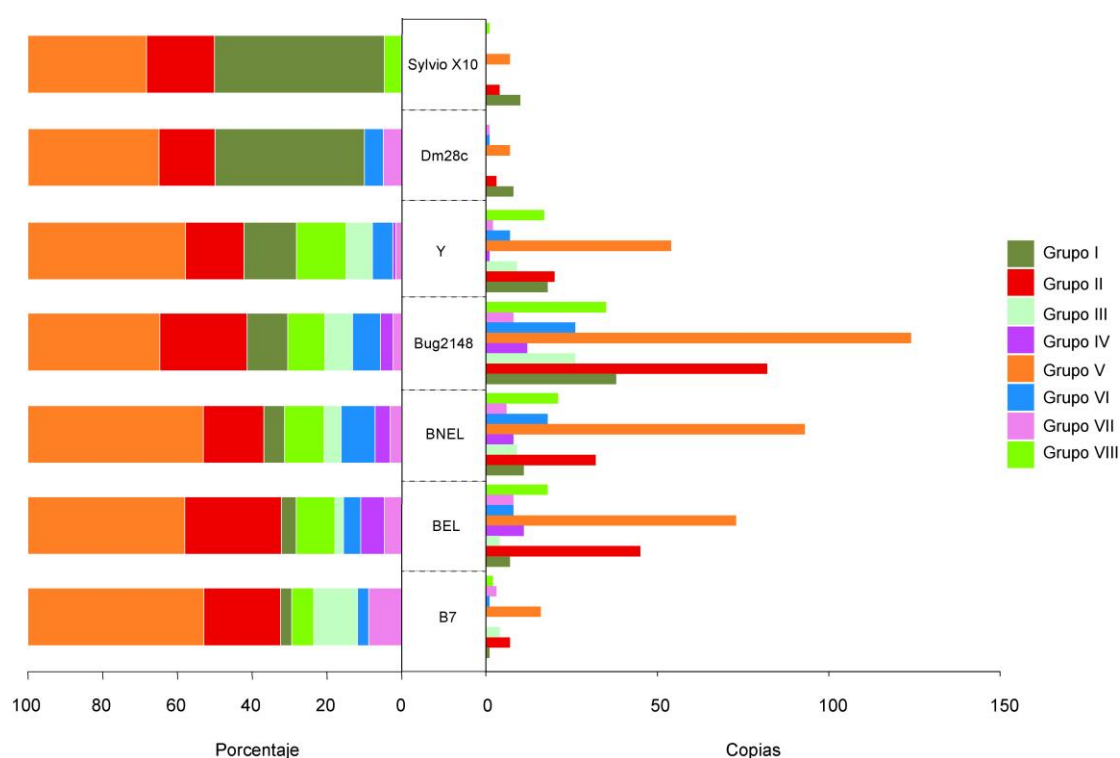
#### **4.1.3 Actividad Trans-sialidasa: subfamilias**

Como se mencionó en la sección anterior, los genes con actividad trans-sialidasa conforman la familia multi-génica más numerosa en *T. cruzi*, con un estimado de al menos 1,400 miembros en cada cepa, y más de la mitad con actividad funcional[30]. La característica principal de esta extensa familia es que comparten el motivo VTVxNVxLYNR[138], y se han descrito como una proteína de membrana presente en trypomastigotes metacíclicos, trypomastigotes de torrente sanguíneo y amastigotes, siendo su principal función la interacción con el huésped[52,139,140]. Su principal función es catalizar la transferencia de moléculas de ácido siálico del huésped a las moléculas receptoras situadas en la superficie del parásito, y por lo tanto se piensa que desempeña un rol imprescindible en la supervivencia del parásito así como el grado de virulencia y efectividad en la infección[141]. Actualmente, los genes codificantes a actividad TS (incluyendo pseudo-genes) son clasificados en 8 grupos, donde cada grupo contiene motivos específicos y que por tanto confieren actividades específicas[52]. A pesar de esto y de la importancia de cada uno de los 8 grupos en procesos específicos de virulencia, la gran mayoría de estas subfamilias no han sido bien definidas y anotadas en los genomas actuales. Por lo tanto, se analizó el contenido de estos 8 grupos en los 2 nuevos genomas ensamblados.

Los genes TS del grupo V que contienen el motivo Asp-Box (SxDxGxTW) es el grupo más abundante, tanto por número de copias como en porcentaje respecto al contenido genético total para la mayoría de cepas incluidas en el análisis excepto para Dm28c y Sylvio X10, ambas de la misma DTU (**Figura 8**). Con hasta 130

copias de esta proteína, Bug2148 y Cl Brener son las cepas con mayor abundancia de este grupo, mientras que, con más de 30, Y es la cepa no híbrida con más genes de este tipo. Este grupo de TSs ha sido asociado a variación antigénica, lo que permite al parásito adaptarse al entorno del huésped[52].

Las TS del grupo II incluye miembros de las glicoproteínas de superficie gp85, TSA-1, gp90, gp82 y ASP-2 cuya actividad está relacionada en procesos de unión al huésped e internalización[52], es el segundo grupo más abundante (por número de copias) en Bug2148 y Cl Brener (124 y 166, respectivamente).



**Figura 8.** Trans-sialidasas en *T. cruzi*. Distribución de los 8 grupos de TS en los genomas actualmente ensamblados. A la izquierda el porcentaje de cada grupo respecto al total de TS, a la derecha el número de copias.

Por otra parte, las TS del grupo I que contienen el mayor número de trans-sialidasas activas, TCNA (*T. cruzi* neuraminidasa) y SAPA por sus siglas en inglés (*Shed Acute-Phase Antigen*) no es tan abundante como se había predicho[142], es posible encontrarlas en todas las cepas, sin embargo las diferencias en número de copias entre cepas es notable. En lugar de las 60-80 copias predichas por cepa, se encontraron anotadas solamente 18 en Cl Brener, 18 en la cepa Y, 38 en Bug2148, 8 y 10 copias en Dm28c y Sylvio X10, respectivamente y 2 en B7.

El grupo VIII es el cuarto más abundante. Las proteínas características de este grupo son la TcS24 y TcS25 que contienen el motivo FRIP (xRxP) y el Asp-box. Las proteínas de este grupo muestran expresión diferencial entre las diferentes formas del parásito, la primera con mayor expresión en trypomastigotes y la segunda en el estado amastigote[52]. Las cepas híbridas contienen el mayor número de estas proteínas, 39 en Cl Brener y 35 en Bug2148 mientras que en las cepas restantes el número es menor, 17 en Y, una en Sylvio X10, ninguna en Dm28c y solo 2 en B7.

El grupo III cuya proteína representativa es la FL-60 y la CRP (*Complement Regulatory Protein*) y que incluye a proteínas complementarias reguladoras e inhibidoras de la convertasa C3, una proteína central en la activación de rutas complemento del sistema inmune, mediante las cuales se elimina a los patógenos del huésped[143] no se encontró en las cepas de la DTU I (Sylvio X10 y Dm28c), pero sí en el resto, siendo las cepas híbridas las más enriquecidas con este grupo. Confirmando que esta podría ser una de las causas por las que algunas cepas tienen mayor capacidad en la evasión del sistema inmune, como se había sugerido anteriormente[144]. Por lo tanto, los resultados obtenidos en este trabajo describen y aumentan la cantidad de información de este grupo genético contribuyendo significativamente en futuros análisis.

El grupo VI, formado por proteínas con los motivos Asp-box y VTVxNVxLYNR se encuentran en cantidades más o menos equitativas (a nivel de porcentaje), excepto en la cepa Sylvio X10 donde no existe ninguna copia anotada.

El grupo IV lo conforman TSs con homología a la proteína más representativa de este grupo, la TsTc13, cuya función específica es desconocida, siendo consideradas como miembros de esta familia simplemente por compartir el motivo conservado VTVxNVxLYNR y agrupación por homología. Curiosamente y de la misma forma que ocurre con las mucinas, no fue posible encontrar genes TS del grupo IV para las cepas clasificadas dentro de la DTU I (Sylvio X10 y Dm28c) y la cepa B7 sugiriendo un nivel de conservación entre estas cepas evolutivamente cercanas[145], sin embargo, en las cepas híbridas el contenido genético es más alto con hasta 15 copias y 3 copias en Y (DTU II). Considerando que este fenómeno podría ser causado por la calidad de los ensamblajes y/o la anotación, no podemos descartar la existencia de este grupo en las cepas de análisis.

Finalmente, las TS del grupo VII caracterizadas por la conservación del motivo xTVxxVxLYNx con la proteína TcS32 como representativa, asociada también en procesos de evasión del sistema inmune, es el grupo menos abundante en todas las cepas. No se encontró ninguna proteína de este tipo en Sylvio X10, solo 1 en Dm28c e Y y 3 en B7.

#### 4.1.4 Estructura cromosómica en *T. cruzi* y diversidad genética

*T. cruzi* posee un genoma complejo a distintos niveles. Plasticidad cromosómica, organización genómica inusual y divergente entre cepas, mecanismos de expresión complejos como transcripción policistrónica, trans-splicing y edición de RNA[9,35,146–149]. Además, en la falta de mecanismos de control del inicio de la transcripción, grupos de genes son regulados post-transcripcionalmente en respuesta a estímulos externos[37]. A día de hoy la organización y mecanismos de regulación de estos grupos es desconocida.

Más del 50% del genoma de este parásito se compone de secuencias repetidas y aunque es considerado un organismo diploide, se ha demostrado gran variabilidad en el número de cromosomas y aneuploidias entre cepas distintas e incluso clones de la misma cepa[133,148,150].

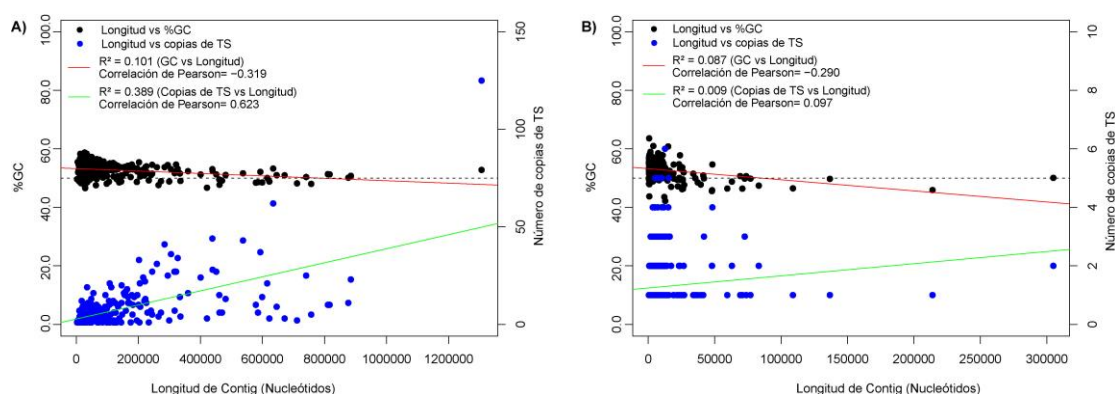
Una de las principales hipótesis de este fenómeno surge de la relación potencial que existe entre la variabilidad y abundancia de las familias genéticas más abundantes en *T. cruzi* con la complejidad genómica de las distintas cepas actualmente estudiadas, relacionadas con la capacidad de evasión de la respuesta inmune en el huésped y de adaptación al ambiente adverso en su ciclo de vida doméstico y selvático.

Se ha demostrado que algunas de las familias genéticas más abundantes en kinetoplastidos, especialmente *T. cruzi*, se localizan en regiones teloméricas y sub-teloméricas y por lo tanto son susceptibles a procesos evolutivos continuos[151]. Estos sitios favorecen la recombinación, expansión del ADN, así como la formación de nuevas variantes genéticas, principalmente se han descrito genes que pertenecen a las familias de DGF, RHS, TS y otras acetil-transferasas ubicados en estas regiones[151–153].

Particularmente, los genes de TSs se localizan en regiones teloméricas, facilitando la generación de nuevas variantes vía recombinación no-homóloga como la gp85 (TS grupo II, la más abundante), siendo este mecanismo el de mayor aceptación sobre la capacidad de evolución del parásito para evadir la respuesta inmunitaria del huésped. Sin embargo, esta importante ventaja evolutiva del parásito, podría ser una de las principales causas de la sub o sobre-estimación de estas familias multi-génicas en los genomas actualmente ensamblados, por las dificultades que supone reconstruir secuencias largas y altamente repetitivas con las tecnologías actualmente disponibles. Los diferentes tipos de repeticiones encontradas en *T. cruzi* son la causa de los altos niveles de fragmentación en los genomas actuales, en

comparación a otros parásitos evolutivamente relacionados como *Leishmania* o *T. brucei*, tanto las familias multi-génicas que representan un alto porcentaje del contenido total, como motivos estructurales altamente repetitivos entre los que se encuentran las repeticiones teloméricas.

En este sentido, y con la hipótesis de que genes expuestos a recombinación y evolución constante contienen elevados niveles de los nucleótidos G y C[154,155], se analizó el contenido de G+C (%GC) en contigs con TS anotadas (incluyendo pseudo-genes) como una medida indirecta de la complejidad en estas regiones. Los resultados muestran que, en ambos genomas, estos contigs con TSs anotadas contienen en general una cantidad más elevada de %GC que se correlaciona de forma negativa con su longitud (correlaciones de Pearson de -0,31 y -0,29 para Bug2148 e Y, respectivamente) en acuerdo con la hipótesis inicial sobre el rol de las repeticiones estructurales asociadas a las TSs (**Figura 9**). Además, como la evidencia actual muestra que la disposición de los genes en kinetoplastidos a lo largo de los cromosomas no sigue un patrón evidente, salvo algunas excepciones donde se han encontrado repeticiones en tándem del mismo gen (principalmente *Leishmania*), los *clusters* (grupos) policistrónicos contienen genes sin relación funcional[153]. Se analizó además la relación que existe entre la longitud de los contigs y el número de copias de TS contenidas, demostrando que actualmente existe correlación positiva principalmente en Bug2148 (0,62), contigs ensamblados con la tecnología PacBio que por sus características (lecturas largas) es capaz de evitar regiones complejas, como las presentes en regiones cromosómicas con TSs abundantes, mientras que en la cepa Y no existe prácticamente correlación debido al nivel de fragmentación, donde el contig más largo apenas alcanza las 300 Kb y con solo 2 genes de TS.

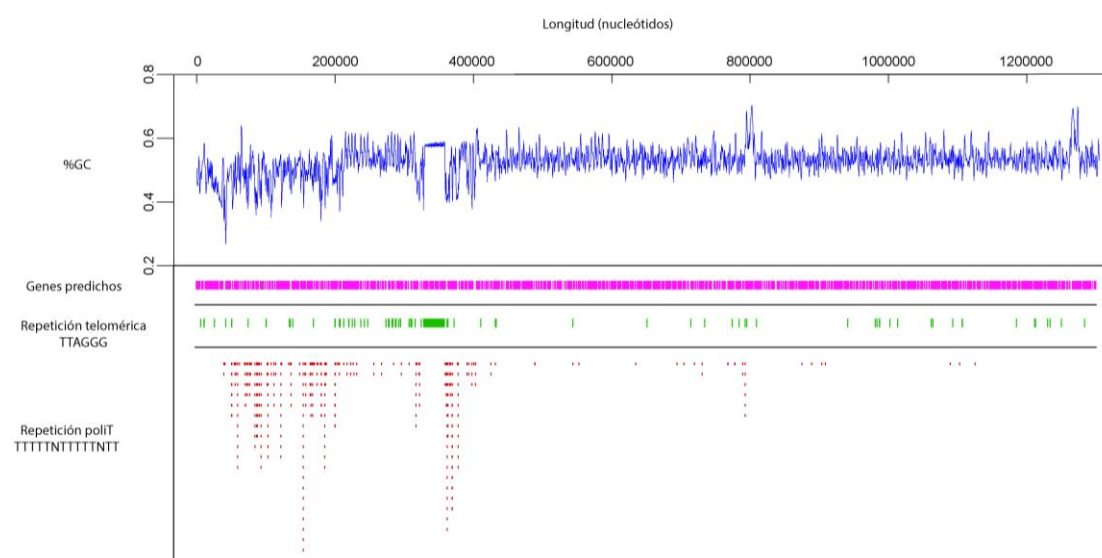


**Figura 9.** Longitud versus %GC (círculos negros), longitud versus número de copias de TS en contigs (círculos azules) de los 2 nuevos genomas. La línea roja indica el



ajuste del modelo lineal y en verde se indica la correlación de ambos parámetros. A) Cepa Bug2148 y B) cepa Y.

Después de este análisis de la complejidad en fragmentos genómicos con TSs, se determinó la localización específica de repeticiones teloméricas (TTAGGG) y estructurales (poly-T: TTTTNTTTTNTT) descritas en *T. cruzi*[156,157], usando como referencia el contig 1 de Bug2148, el más grande ensamblado *de novo* hasta la fecha en *T. cruzi* (1,3 Mb), con más genes TSs anotados (120) y que además se corresponde con uno de los cromosomas más grandes en esta especie (**Figura 10**). Este análisis demostró que el %GC en el genoma del parásito es dependiente de estas repeticiones, y por lo tanto podrían ser la causa del colapso y fragmentación de la mayoría de genomas en proyectos *in silico*, causando la sub y sobre-estimación de las familias más abundantes localizadas en estas regiones[157], principalmente TSs y proteínas DGF. Pero adicionalmente nos permite hipotetizar sobre el resto de familias genéticas presentes y su predisposición a evolución constante mediante procesos de duplicación, mutación y recombinación como se había sugerido anteriormente[157].

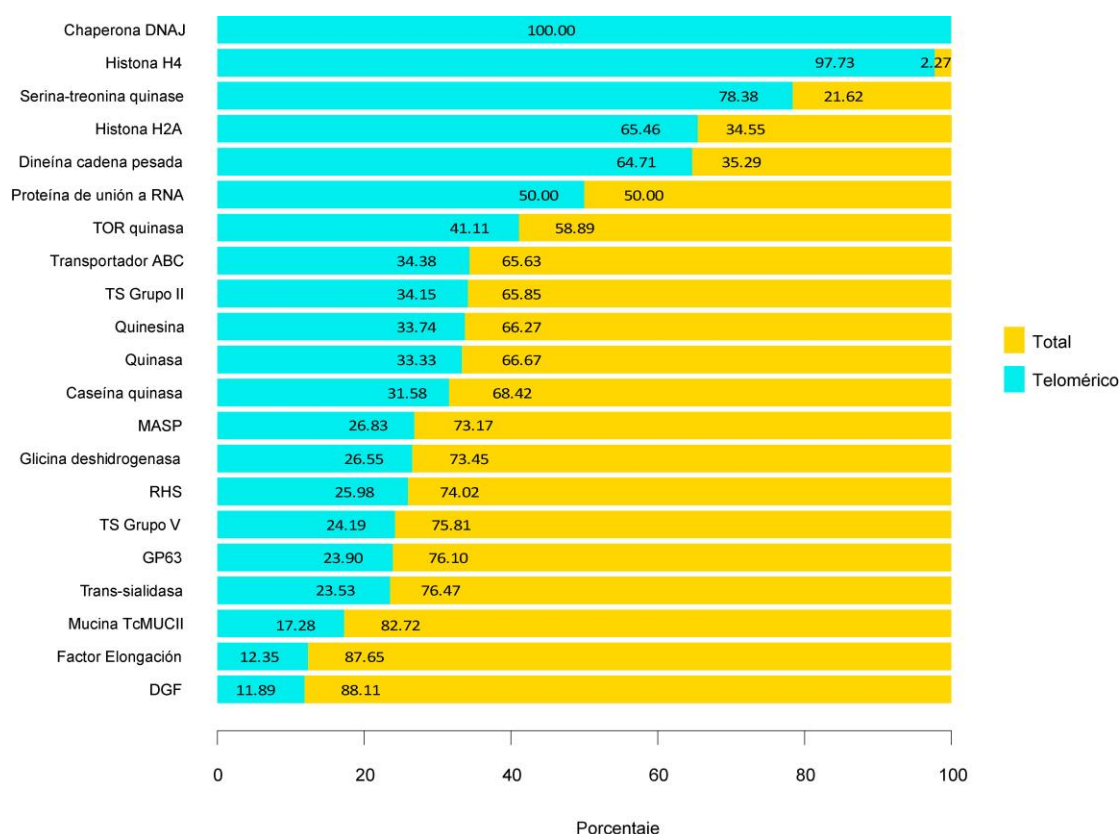


**Figura 10.** Distribución del %GC en el contig 1 de Bug2148 y su relación con la distribución de repeticiones teloméricas y estructurales descritas en *T. cruzi*.

Adicionalmente, se encontraron repeticiones en tándem de genes codificantes para las proteínas caseína quinasa la región con más repeticiones teloméricas. Se trata de una de las proteínas con actividad quinasa más importante de función celular, común y conservada en la mayoría de eucariotas que pertenece a la familia multipotencial Serina/Treonina con capacidad de fosforilar una amplia variedad de proteínas[158]. Sin embargo, su rol específico en kinetoplastidos no se ha descrito,

impidiendo hipotetizar sobre las razones de su localización y potenciales consecuencias.

Asimismo, se analizó la composición genética en contigs con 20 o más repeticiones teloméricas en tándem como marcador de estas regiones cromosómicas en el genoma de Bug2148 (**Figura 11**). Trabajos anteriores sugieren que aproximadamente 9% de los genes con actividad TS, 12% de genes DGF y 19% de RHS se localizan en estos sitios[152], sin embargo, los resultados obtenidos demuestran que estos porcentajes podrían ser mayores y que deben incluirse algunas otras familias genéticas, especialmente quinasas, MASP, mucinas y proteínas relacionadas en el proceso de transcripción y traducción como las histonas, las proteínas de unión a RNA y los factores de elongación. Además de la Gp63, la sexta proteína más abundante según este trabajo, y de acuerdo con lo esperado, el porcentaje de TS en estos sitios podría ser más del doble de lo predicho anteriormente, al menos para cepas híbridas. Además, al menos una cuarta parte de las TS más abundantes del grupo V y II se encuentran también en estas regiones.



**Figura 11.** Porcentaje de genes anotados en contigs con 20 o más repeticiones teloméricas en tándem (Azul) respecto al total (amarillo) en la cepa Bug2148.

Al menos un 25% de los genes MASP se encuentran también en estas regiones expuestas a evolución constante, confirmando previas suposiciones sobre la divergencia de esta abundante familia y su importante papel en la genómica y transcriptómica del parásito[159].

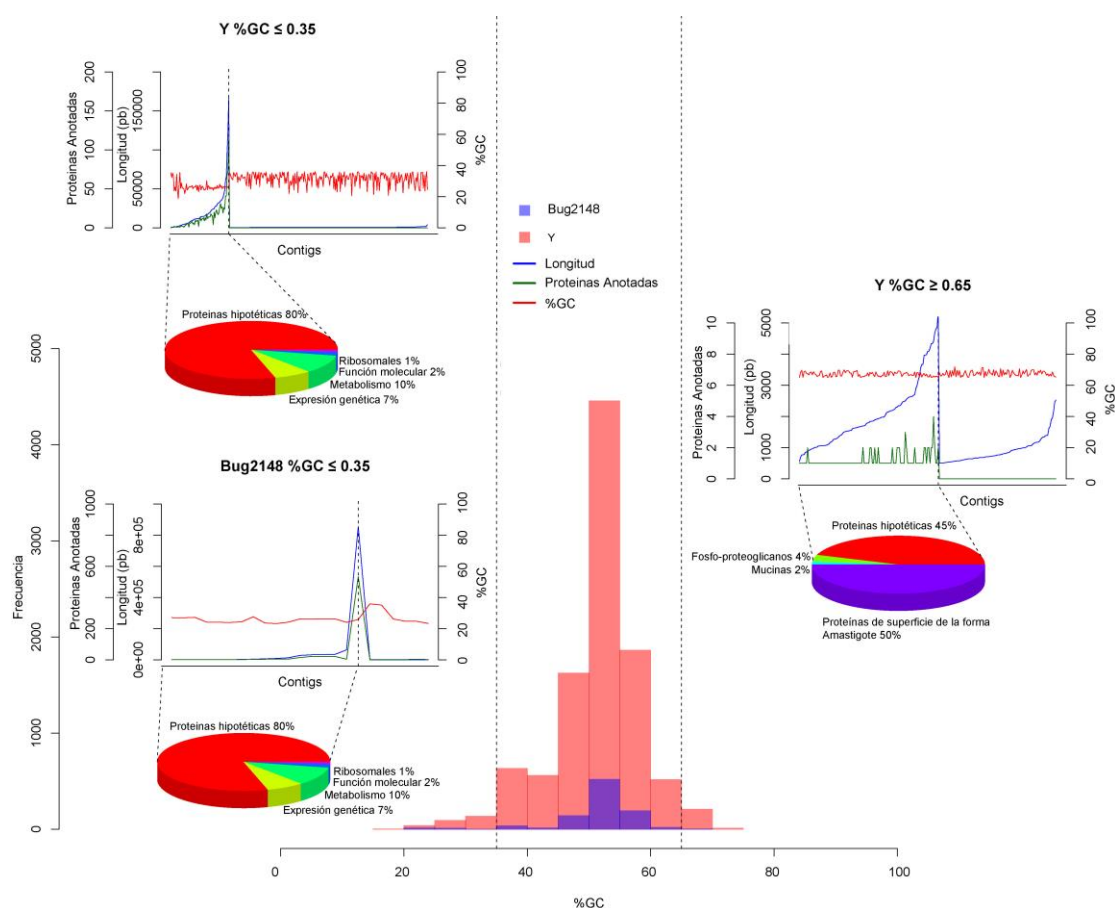
El ADN en *T. cruzi* se organiza en nucleosomas, conformado por histonas canónicas H2A, H2B, H1, H3, y H4 y algunas variantes (H2A.Z, H2B.V, H3.V)[160]. Además, las histonas en kinetoplastidos se caracterizan por algunas peculiaridades, principalmente, existe más de un gen codificante para cada proteína (canónica o variantes) y muestran altos niveles de divergencia a nivel de secuencia, mientras que en la mayoría de eucariotas son proteínas altamente conservadas[160].

Asombrosamente, los resultados muestran que otras proteínas consideradas como menos abundantes en *T. cruzi* (por número de copias) como H4, H2A, chaperona DNAJ, factores de elongación, y proteínas de unión a RNA se localizan también en estas regiones cromosómicas complejas, lo que nos permite hipotetizar sobre dos posibles consecuencias, por un lado, la predisposición a evolución constante como sucede en las familias multi-génicas de proteínas de superficie y por otra parte, el rol fundamental de genes teloméricos en la cromatina telomérica, su replicación, pero principalmente la regulación de transcritos no teloméricos mediante su actividad en procesos de transcripción, procesos post-transcripcionales y en la traducción[161]. Por otra parte, y en acuerdo con trabajos anteriores, mientras que las familias de proteínas de choque térmico principalmente Hsp70 y Hsp40 muestran niveles de conservación entre el 60-90%, las DNAJ presentes en regiones teloméricas muestran niveles de similitud entre el 15-60%, potencialmente como resultado de su mayor susceptibilidad a procesos evolutivos[162].

En el caso de la cepa Y, solamente se encontraron 11 contigs con 20 o más repeticiones teloméricas repetidas, y de longitud máxima de 3 kb. Las proteínas anotadas en estos contigs se corresponden en su mayoría con proteínas hipotéticas, 3 proteínas con función kinasa y 1 *TS-like*. Como se mencionó anteriormente, el genoma de *T. cruzi* se compone por tres clases de DNA: 1.- secuencia codificante (genes de copia única conservados entre cepas y especies), 2.- secuencia codificante de familias multi-génicas con similitud variable (principalmente proteínas de superficie y asociadas a virulencia) y 3.- secuencias no codificantes y repetitivas (micro y mini-satélites y repeticiones cortas y largas intercaladas en zonas codificantes). Estos dos últimos tipos de secuencia constituyen cerca del 50% de la secuencia genómica total, afectando gravemente proyectos de secuenciación con tecnologías actuales basadas en la generación de secuencias cortas (hasta 300 nucleótidos) como la empleada en este trabajo para la cepa Y. Por lo tanto, los

genomas ensamblados mediante esta tecnología (Illumina), o incluso con ensamblajes híbridos (mezclando dos tecnologías de características diferentes, p. ej Illumina y Roche 454) pero de baja cobertura resultan con altos niveles de fragmentación, impidiendo la reconstrucción de la estructura genética y consecuentes hipótesis, como en el caso de la cepa Bug2148 en estas regiones de alta complejidad lingüística.

Sin embargo, como el %GC en algunas especies eucariotas se correlaciona con características moleculares potencialmente interesantes como, distribución, elementos transponibles, metilación, niveles de expresión y los genes en regiones de %GC alto contienen también niveles elevados de GC en comparación con regiones no codificantes, lo que supone una ventaja selectiva por la eficiencia de su procesamiento[163]. Se analizaron, en ambos genomas completos, los contigs con los más altos y bajos niveles de %GC (**Figura 12**).



**Figura 12.** Distribución de %GC y anotación genética en los nuevos genomas de Y y Bug2148. El histograma central muestra la distribución completa de GC en ambos genomas, a los extremos izquierdo y derecho se muestran las anotaciones y longitud de los contigs con los niveles más bajos ( $\leq 35$ ) y altos ( $\geq 65$ ) de %GC.

En la cepa Y, se encontraron un total de 316 contigs con los %GC más bajos y 71 con al menos una anotación (cerca del 22%), mientras que 219 contigs contienen un %GC superior al 65% y más de la mitad contienen algún gen anotado (119). En el caso de Bug2148, 23 contigs contienen %GC bajo, 17 de ellos con al menos 1 anotación (74%) y solamente 2 contigs con %GC alto y conteniendo genes de función hipotética. En ambos casos, los contigs de %GC bajo se corresponden a regiones cromosómicas con genes de función constitutiva, tales como genes ribosomales, asociados al metabolismo y expresión génica (además de genes con proteínas hipotéticas). En el caso de Y, tal y como se esperaba por las desventajas de la tecnología empleada, la mayoría de los contigs no contienen proteínas anotadas. Sorpresivamente, los genes en contigs de la cepa Y con los %GC más altos contienen proteínas de superficie, principalmente asociadas a la forma amastigote, mucinas y asociadas a fosfo-proteoglicanos. Estos resultados sugieren, que los niveles de %GC podrían conferir una ventaja adaptativa a una forma específica del parásito, sujeto a estrés oxidativo y/o choque térmico causada por el huésped en respuesta a la infección, tal y como se ha propuesto con otros grupos proteicos en otras especies[154,155].

#### **4.1.5 Genes de copia única en *T. cruzi***

Con los resultados del proceso de agrupación por similitud de secuencia (MCL) y los resultados de Blastp, se determinaron un total de 400 y 183 genes de copia única con función conocida en los nuevos genomas de las cepas Bug2148 e Y, respectivamente. Además, cada gen fue asociado a una función molecular mediante ontología genética (*Gene Ontology: GO*) (**Anexos, apéndice A, Figura Suplementaria 1**).

En la cepa Y se obtuvo una cantidad relativamente menor de genes de copia única de los 215 estimados en esta especie[164], mientras que como se esperaba para la cepa híbrida, los obtenidos en la cepa Bug2148 constituyen casi el doble de esta cantidad, representando así un avance sustancial en el conocimiento de la biología molecular del parásito. Tal como se esperaba, los genes de copia única en la cepa Y se encuentran conservados casi en su totalidad en la cepa Bug2148, y esta a su vez conserva una variabilidad mayor de este tipo de genes.

Las grandes diferencias observadas en el grado de completitud de ambos genomas respecto a la tecnología empleada nos permiten confirmar la calidad del genoma de la cepa Bug2148, sin embargo, es precisamente esta diferencia de calidad que nos permite hipotetizar sobre la importancia de genes de copia única exclusivos de la

cepa Y. Los 2 genes de copia única exclusivos de la cepa Y se corresponden a una proteína de glicosilación ligada a asparagina (ALG12) que pertenece al grupo de procesos biosintéticos derivados de carbohidratos, y una proteína de biosíntesis cofactor de molibdeno o Moco por sus siglas en inglés (*Molybdenum cofactor biosynthesis protein*), perteneciente al grupo prostético metabólico. Poco se sabe de la ALG12 en *T. cruzi*, se trata de una proteína conservada en eucariotas, cuya función es la glicosilación (en asparaginas) de proteínas nacientes en el retículo endoplásmico[165]. Por otra parte, Moco es un cofactor de molybdopterina de la xantina oxidasa, DMSO reductasa, sulfito oxidasa, nitrato reductasa y otras oxidasas implicadas en el metabolismo de purinas[166]. Interesantemente, un estudio de Jose Luis Ávila y Ángela Ávila[167] demostró en modelos *in vivo* que el uso de inhibidores de la xantina oxidasa, como el alopurinol produce actividad tripanostática, protegiendo al huésped de la infección aguda.

Aunque se debe tener en cuenta que estas diferencias podrían ser originadas por las diferencias en las tecnologías de secuenciación empleadas, tal como ocurre en otros grupos de genes más abundantes, estos genes pueden constituir un objetivo relevante en futuras investigaciones debido a la potencial importancia de su existencia.

La identificación de los contenidos genéticos específicos de cada cepa, y específicamente de los genes de copia única puede constituir una estrategia interesante para entender el comportamiento diferencial del parásito, tanto en su ciclo de vida como en el proceso de infección y las diferencias de patogenicidad entre cepas. Además, estos genes pueden constituir un blanco más específico y adecuado para técnicas de edición genética emergentes como el CRISPR/CAS9, en organismos donde un alto porcentaje de la composición genética total se constituye de genes multicopia.

## **4.2 Transcriptómica de la forma infectiva de la cepa Sylvio X10 Y refinamiento de la secuencia genómica**

### **4.2.1 Refinamiento de la secuencia genómica, ensamblaje del transcriptoma y cuantificación**

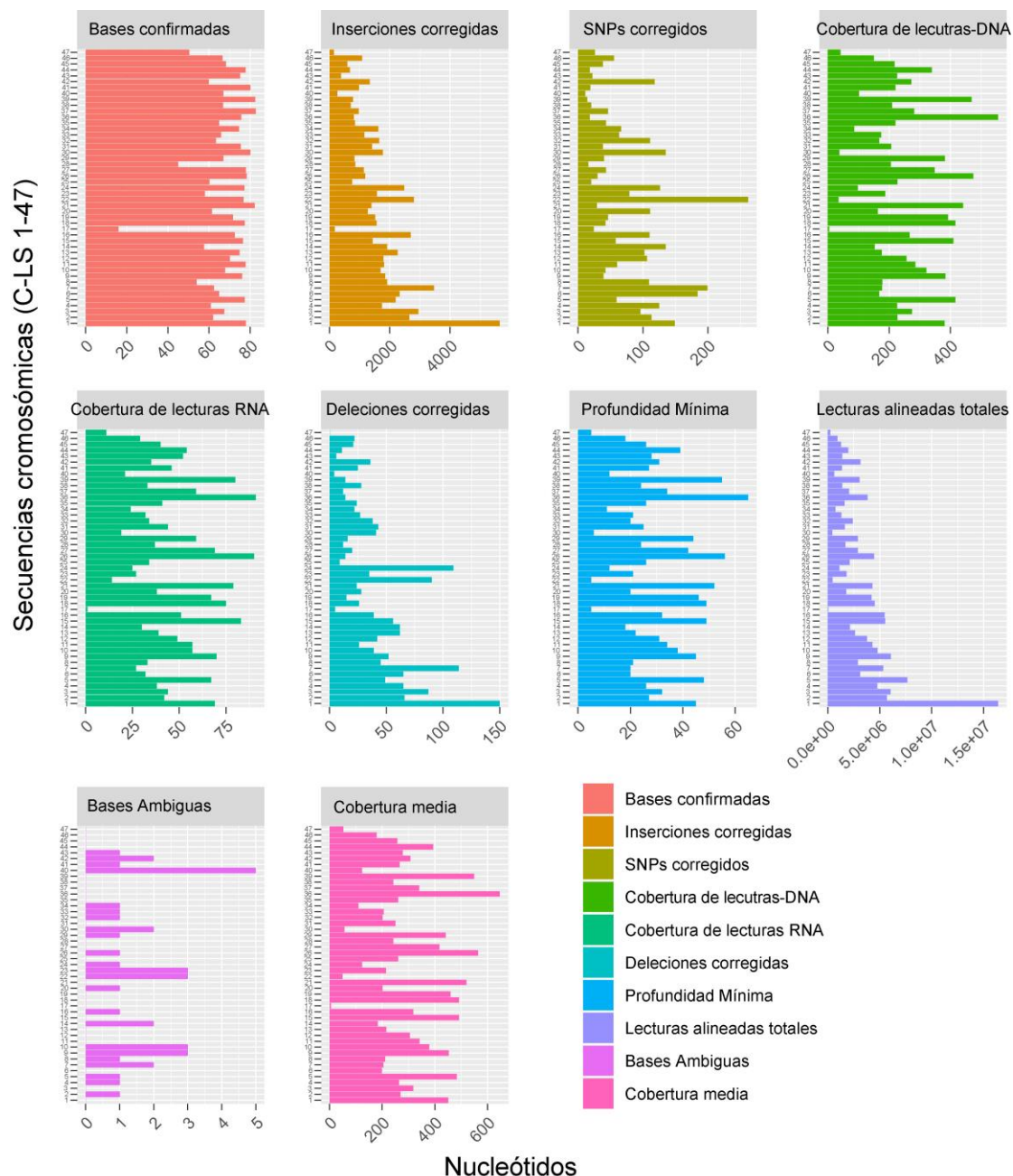
Se secuenció el ARN de tripomastigotes (cepa Sylvio X10) con la tecnología Illumina en la plataforma MiSeq, después de una selección de ARNm por polyA+, y se obtuvieron un total de 17,332,912 secuencias pareadas (*paired-end*) con longitud

promedio de 75 nucleótidos. Estos datos crudos de secuenciación se depositaron en la base de datos SRA (*Sequencing Read Archive*) del NCBI con el número de acceso PRJNA546488.

Como primer paso, las lecturas obtenidas fueron recortadas y filtradas, estableciendo como longitud máxima y mínima 100 y 50 nucleótidos, respectivamente, la calidad mínima permitida fue de 25 (escala *phred*). Estas lecturas filtradas fueron mapeadas al genoma más reciente de SylvioX10 (01/Noviembre/2018) depositado en el Trytrip (<http://tritrypdb.org/>, "TcruziSylvioX10-1) con Bowtie2 permitiendo hasta tres *mismatches*. Considerando que el 81.75% de las lecturas mapearon satisfactoriamente al genoma, se usó este ensamblaje como referencia *bona fide*.

Este genoma de referencia está compuesto por 47 *scaffolds* (secuencias largas simulando cromosomas completos) o C-LS por sus siglas en inglés (*Chromosome-like sequences*) y no se corresponden a una versión final completa del genoma de esta cepa. Por lo tanto, es necesario resaltar que las lecturas restantes no alineadas obtenidas del RNAseq pueden corresponderse a estas regiones cromosómicas ausentes o no resueltas (gaps entre contigs). Por consiguiente, estos datos de secuenciación fueron usados como información adicional para mejorar la secuencia genómica ensamblada, mediante la corrección de mutaciones (SNPs: *Single Nucleotide Polymorphisms*), deleciones, inserciones etc., con el software Pilon. Esta herramienta bioinformática permite la corrección de genomas haploides y diploides incompletos usando lecturas pareadas de NGS (principalmente Illumina). Adicionalmente se usaron lecturas de DNaseq de esta misma cepa depositadas en las bases de datos del NCBI bajo con el número de acceso PRJNA395140.

Los resultados de este proceso de corrección de secuencia se resumen en la **Figura 13**. 68% de la secuencia genómica total fue confirmada (sin contar los gaps), correspondiéndose a aproximadamente 28 Mb de información.



**Figura 13.** Genome improvement summary. Read depth mapping, coverage, and changes applied to the previous genomic sequence of Sylvio X10.

La mayor parte de la secuencia corregida se corresponde con regiones codificantes, determinada por las lecturas de RNAseq. En total se corrigieron un total de 3.415 SNPs, 39 bases ambiguas, 72.887 inserciones y 1.810 deleciones. 18 de los 47 C-LS decrecieron en longitud. Dentro de los C-LS que más acortaron su secuencia total destacan, con hasta 23 kb el C-LS 4, el C-LS 1 (el más largo antes de la corrección) decreció 7,2 kb, el segundo más largo 12,8 kb, mientras que el más corto (C-LS 47) también decreció 7,2 kb.



En total, 133,35 kb de inserciones fueron recortadas del ensamblaje inicial, mientras que fue posible añadir 43,66 kb de nueva información (por deleciones corregidas), lo que supone un total de 177,02 kb de correcciones a lo largo de esta nueva versión del genoma.

Cabe resaltar que los C-LS 17 y 47 obtuvieron la menor cobertura de secuenciación por DNA y RNA, y contienen además la más alta relación de gaps por nucleótido ensamblado, sugiriendo que estas secuencias podrían corresponderse con regiones estructurales complejas o simplemente ensamblajes espurios.

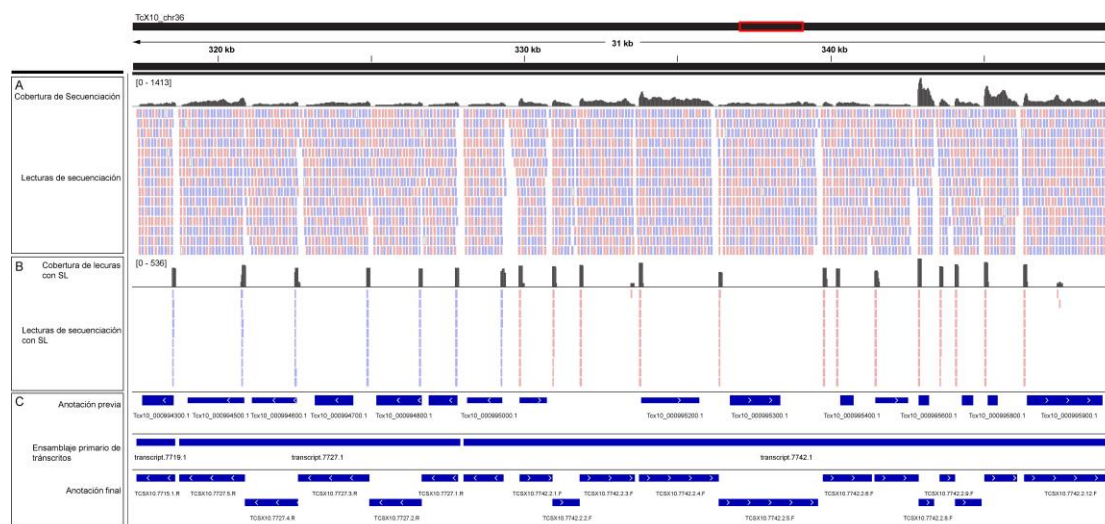
Después de este proceso de corrección, las lecturas filtradas fueron mapeadas nuevamente a la nueva versión del genoma, obteniendo un total de 82,35% de alineamiento y se continuó con el análisis.

Como se describió anteriormente, los genes en tripanosomatidos no contienen promotores ni existe algún mecanismo específico por el cual se controle la transcripción genética. En contraste, grupos de genes se transcriben al mismo tiempo en RNAs largos (de hasta varias decenas de kb) en lo que se conoce como unidades de transcripción poli-cistrónica, o PTUs por sus siglas en inglés (*Polycistronic Transcription Units*), requiriendo a continuación de procesos de maduración adicionales conocidos como trans-splicing y poli-adenilación, mediante los cuales se generan los mRNAs individuales y maduros[146,147,168]. Con esta información como referencia, el transcriptoma completo fue obtenido en dos pasos, primero, usando el software StringTie se realizó un ensamblaje primario de las secuencias, obteniendo un total de 9.108 transcritos que se corresponden con estimaciones previas[41,169], pero menos del 50% del repertorio genético total estimado en este trabajo para una cepa ancestral (20.058)[170] o incluso menos que los genes anotados en la versión anterior del genoma de referencia (20.619 incluyendo pseudogenes).

No se puede descartar la presencia de PTUs estables y sin procesar en el transcriptoma del parásito como las causantes de este ensamblaje primario, tal y como se ha descrito en otras especies relacionadas[171], sin embargo, fue posible detectar potenciales mRNAs individuales dentro de las PTUs ensambladas mostrando claramente diferencias de abundancia (**Figura 14A**). Sorpresivamente, la mayoría de estos potenciales mRNAs individuales no coincidió con genes anotados en la primera versión del genoma.

Como segundo paso, para definir los transcritos individuales contenidos en las PTUs del ensamblaje primario, se usaron como referencia la secuencia de 39 nucleótidos (SL) y cadenas poli-A en los extremos 5' y 3' UTR (*Untranslated Region*) de cada

transcrito descritas anteriormente[172–174], así como las coordenadas de las lecturas que las contenían, como se describe en el trabajo publicado por Rastrojo y col. (2013)[39].



**Figura 14.** Ensamblaje y anotación del transcriptoma de *T. cruzi*. Se muestra una región de 31 kb en el C-LS 2, **A:** cobertura de secuenciación (RNAseq) y lecturas alineadas, **B:** cobertura y lecturas que contienen la secuencia SL, **C:** anotaciones, anotación previa, ensamblaje primario y anotación del transcriptoma final.

Se encontraron 1.083.343 lecturas conteniendo la secuencia SL, al no formar parte de las UTRs de cada gen, se recortó y se volvieron a mapear estas lecturas al genoma de referencia (**Figura 14B**), obteniendo mediante las coordenadas del alineamiento, la posición exacta de su inserción (se identificaron 42.700 sitios de inserción de SL) y el inicio de la 5' UTR. De la misma forma, se identificaron 85.452 lecturas conteniendo cadenas poli-A (14.512 sitios de inserción) y el extremo 3' UTR. Tal como se ha descrito en *L. major*, la ausencia de secuencias conteniendo SL en dirección contraria al sentido de la ORF, confirma que el mecanismo de transplicing ocurre solamente en el mismo sentido de la transcripción de la PTU[39]. Con esta información, la mayoría de las PTUs ensambladas en el primer paso fueron finalmente divididas en 18.666 transcritos individuales. Del total, 1.172 fueron anotados como policistrónicos al contener más de una ORF anotada pero que no fueron divididos por la ausencia de lecturas con SL o poli-A. 872 fueron anotados como transcritos truncados, ya que la cobertura es menor al 50% de la longitud de la ORF anotada. Aunque estos transcritos policistrónicos y truncados necesitan ser confirmados en futuros análisis de RNAseq o DNaseq, en prácticamente el 100% de

los casos se trata simplemente de transcritos con baja expresión (al menos en el estado tripomastigote).

Por otra parte, considerando que la secuencia previa del genoma fue modificada, la nueva versión fue re-anotada usando el Software Companion y la evidencia transcriptómica obtenida en este trabajo.

Las lecturas restantes no alineadas a la nueva versión del genoma o a secuencias conocidas de mini y maxi-círculos (aproximadamente 3 millones de lecturas pareadas) fueron ensambladas *de novo* con el software rnaSPAdes, con el objetivo de reconstruir la porción del transcriptoma no ensamblado por la falta de referencia genómica. Los transcritos potenciales ensamblados fueron filtrados por cobertura de secuenciación ( $\leq 10$ ), longitud ( $\geq 1$  kb), redundancia de secuencia (similitud  $\geq 90\%$ ) y ORFs contenidas (transcritos con más de una ORF de más de 50 aa fueron descartados). Se obtuvieron un total de 4.311 transcritos adicionales (9,62 Mb de información extra), su abundancia relativa fue igualmente calculada con el software StringTie y fueron añadidos al transcriptoma inicial, obteniendo un total de 22.977 transcritos totales, individuales y maduros.

Estos transcritos individuales fueron nombrados usando la misma nomenclatura que en otros tripanosomatidos: iniciando por la especie "TC" (*T. cruzi*) seguida de la cepa SX10 (SylvioX10), el número del policistrón al que pertenece el transcrito, el monocistrón (o número de transcrito) y el sentido "F" (*Forward*), "R" (*Reverse*) o X (No definido), p.ej. TCSX10.1.1.F. Para los transcritos ensamblados *de novo*, el número de policistrón fue remplazado por "ND" (del inglés *non-defined*) y el número de monocistrón por el número de contig (asignado por defecto por rnaSPAdes, ordenando los transcritos por longitud en orden descendente) y el sentido definido por el sentido de la ORF contenida, p.ej. TCSX10\_ND.1.F.

En conjunto, los resultados demuestran que el 79,95% de la secuencia genómica se corresponde a secuencia codificante y, por lo tanto, aproximadamente el 20% es secuencia estructural no codificante. Este resultado es de especial relevancia, ya que la versión anterior de este genoma contenía anotada solamente el 37.73% de su secuencia (incluidos pseudogenes). De este modo, este resultado duplica la estimación anterior y mejora dramáticamente la anotación anterior (**Figura Suplementaria 2**). Además, con base a la longitud total de los transcritos *de novo*, podemos suponer que la longitud total del genoma haploide de esta cepa podría ser mayor que lo estimado (al menos 51 Mb).

17.334 transcritos tienen un FPKM superior o igual a 10 y el cuartil 3 incluye valores de hasta FPKM= 66, mientras que 59 de los transcritos contienen la máxima

abundancia con valores fuera de la distribución normal (valores atípicos) por encima de 1000 (**Figura suplementaria 3**).

12.146 de los transcritos incluidos en la referencia genómica (65%) y 4.197 (97%) de los transcritos ensamblados *de novo* fueron identificados mediante Blastx (incluidos genes de función hipotética conservada). Los transcritos cuya ORF no fue identificada por Blast en las bases de datos utilizadas, fueron anotados como “*Not Found*” en el archivo de anotaciones.

Los 59 transcritos más abundantes (FPKM mayor o igual a 1.000) así como la función del gen contenido se detallan en la **Tabla 4**. Sorprendentemente, 83% (49) de los 59 transcritos más abundantes se corresponde a los ensamblados *de novo*, y que, por lo tanto, no están ensamblados en la secuencia genómica de esta cepa.

El transcrito más abundante (TCSX10\_ND.1840.F) contiene al gen codificante EF-1- $\alpha$  (*elongation factor 1-alpha*). Es un gen altamente conservado en eucariotas, cuya función principal función de este gen es la unión a GTP que a su vez interacciona con aminoacil-ARNt durante la formación de polisomas[175], y por lo tanto ha sido involucrado en gran variedad de procesos celulares, tales como, metabolismo, organización del citoesqueleto, apoptosis y principalmente como potencial regulador de la expresión génica en *T. cruzi*[176]. Curiosamente, aunque se estima la existencia de al menos 11 copias de este gen repetidas en tándem en el genoma de *T. cruzi*[176], fue posible determinar un segundo transcrito de este gen (TCSX10\_ND.7486.F) dentro del grupo de los más abundantes. La variabilidad en la longitud y similitud de la ORF en ambos transcritos permitió su correcta identificación y cuantificación por ensamblaje *de novo*, y por otro lado confirma la incapacidad de los secuenciadores y ensambladores actuales en regiones codificantes de alta complejidad.

El segundo y quinto transcritos más abundantes corresponden a genes de las glicoproteínas de superficie más abundantes en esta especie, y que además han sido relacionadas con procesos de invasión celular, virulencia y supervivencia[128] (principalmente en cepas de la DTU I), las mucinas. De acuerdo a lo esperado por su importante papel en la regulación de la expresión génica de kinetoplastidos, tales como *Leishmania* (promastigotes) donde constituyen hasta el 2,1% de la proteína total en condiciones sin ningún tipo de estrés[177], 4 de los más abundantes mRNAs contienen ORFs de proteínas de choque térmico o *Hsp*, por sus siglas en inglés (*Heat shock proteins*).

Las proteínas de unión a ARN o RBPs (*RNA binding proteins*) desempeñan también una función fundamental en procesos post-transcripcionales mediante la

estabilización de los ARNm y la biogénesis de los ribosomas[178,179], y por lo tanto han sido objeto de estudio durante los diferentes estados del parásito y su ciclo de infección, sin embargo, se encontró solamente un transcrito (dentro de los 59 con mayor abundancia) que codifica esta función (TCSX10\_ND.7601.F, FPKM = 2301,46), pero que de manera interesante contiene el motivo *RRM3* (*RNA recognition motif 3*) de tipo I específico de las PABPs (*polyA-binding proteins*) conservado en proteínas que se unen de forma específica a las cadenas poli-A de los ARNm de eucariotas[180]. Por lo tanto, este transcrito podría constituir un blanco terapéutico interesante al menos en la forma tripomastigote, que merece análisis más específicos y profundos.

Por otra parte, 17 ARNm de esta lista contienen transcritos que codifican proteínas ribosomales (9 para 60S, 5 de 40S y una para S25, L24 y L37). Estos resultados son de especial importancia, ya que por primera vez en *T. cruzi* se describen los niveles y tipos de este tipo de ARN en la forma tripomastigote, que a lo largo de los años han constituido las dianas farmacológicas más importantes. Se ha demostrado que pequeñas diferencias en la composición ribosómica de tripanosomatidos pueden servir como objetivos específicos de fármacos contra la enfermedad de Chagas[181–183].

Finalmente, los transcritos restantes más abundantes corresponden a dos de las principales histonas canónicas H2B y H2A (TCSX10\_ND.4352.F y TCSX10\_ND.4559.F, respectivamente), entre otros relacionados con el metabolismo.

**Tabla 4.** Se detallan los 59 transcritos más abundantes en el transcriptoma de *T. cruzi*.

Transcript Name	FPKM	Blastx-Hit
TCSX10_ND.1840.F	7377,97	Elongation factor 1-alpha (EF-1-alpha)
TCSX10_ND.7941.F	4776,96	Mucin-like glycoprotein
TCSX10_ND.3154.F	4669,15	D-isomer specific 2-hydroxyacid dehydrogenase-protein
TCSX10_ND.1768.F	4298,94	Heat shock protein 85
TCSX10_ND.5081.F	3981,67	Mucin-like glycoprotein
TCSX10_ND.4436.F	3773,12	Cystathionine beta-synthase
TCSX10_ND.6083.F	3120,83	Heat shock protein 70 (HSP70)
TCSX10_ND.4352.F	2740,59	Histone H2B
TCSX10.6560.6.R	2501,74	Hypothetical protein
TCSX10_ND.5435.F	2434,05	L-threonine 3-dehydrogenase
TCSX10_ND.5069.F	2373,31	Trans-sialidase

Transcript Name	FPKM	Blastx-Hit
TCSX10_ND.7601.F	2301,47	RNA-binding protein
TCSX10_ND.2296.F	1981,37	Serine carboxypeptidase (CBP1)
TCSX10_ND.6353.F	1942,05	60S ribosomal protein L10a
TCSX10.8474.1.R	1921,65	Mucin-like glycoprotein
TCSX10_ND.466.F	1916,11	Amino acid transporter
TCSX10_ND.6832.F	1890,20	Glutamamyl carboxypeptidase
TCSX10_ND.364.F	1838,19	Hexose transporter
TCSX10_ND.3852.F	1784,70	Putative helicase
TCSX10.4107.1.R	1619,73	40S ribosomal protein S33
TCSX10_ND.3412.F	1616,62	Tyrosine aminotransferase
TCSX10_ND.5699.F	1611,10	Hypothetical protein
TCSX10_ND.3416.F	1605,96	Tyrosine aminotransferase
TCSX10.3691.1.F	1570,57	60S ribosomal protein L44
TCSX10_ND.7216.F	1551,15	Hypothetical protein
TCSX10_ND.1981.F	1549,12	40S ribosomal protein S21
TCSX10_ND.6160.F	1542,95	Sterol 24-c-methyltransferase
TCSX10_ND.7486.F	1431,42	Elongation factor 2
TCSX10_ND.5213.F	1429,50	Ubiquitin/ribosomal protein S27a
TCSX10_ND.7089.F	1428,85	60S acidic ribosomal protein P0
TCSX10_ND.6956.F	1373,65	60S ribosomal protein L5
TCSX10_ND.7682.F	1360,50	60S ribosomal protein L2
TCSX10_ND.326.F	1352,74	Hexose transporter
TCSX10_ND.8140.F	1352,57	Putative mucin TcSMUGS
TCSX10_ND.7552.F	1341,34	40S ribosomal protein S4
TCSX10_ND.3709.F	1316,95	Cysteine peptidase
TCSX10_ND.8348.F	1277,92	60S ribosomal protein L18
TCSX10_ND.2638.F	1255,85	Glycosomal phosphoenolpyruvate carboxykinase
TCSX10_ND.1562.F	1255,38	Glycosomal phosphoenolpyruvate carboxykinase
TCSX10.6382.2.F	1236,96	60S acidic ribosomal protein P2 beta (H6.4)
TCSX10_ND.6305.F	1221,18	Tryparedoxin peroxidase
TCSX10.5222.9.F	1216,11	Kinetoplastid membrane protein KMP-11
TCSX10_ND.8087.F	1215,93	Ribosomal protein S25
TCSX10_ND.5427.F	1163,64	Heat shock 70 kDa protein
TCSX10_ND.220.F	1162,89	P-type H <sup>+</sup> -ATPase
TCSX10_ND.4125.F	1154,22	90 kDa surface protein
TCSX10_ND.5464.F	1145,67	Trans-sialidase
TCSX10_ND.4595.F	1143,34	Enolase
TCSX10_ND.4559.F	1130,09	Histone H2A

Transcript Name	FPKM	Blastx-Hit
TCSX10.457.12.R	1127,90	Ribosomal proteins L37
TCSX10_ND.5709.F	1072,75	Hypothetical protein
TCSX10_ND.4984.F	1067,71	ADP,ATP carrier protein 1
TCSX10_ND.8347.F	1056,89	Trans-sialidase
TCSX10_ND.6727.F	1056,79	Heat shock 70 kDa protein
TCSX10.5300.1.R	1026,09	40S ribosomal protein S11
TCSX10.5088.14.F	1022,97	Ribosomal protein L24
TCSX10_ND.8088.F	1009,79	40S ribosomal protein SA
TCSX10.1814.2.R	1006,31	60S ribosomal protein L17
TCSX10_ND.7034.F	1000,45	60S ribosomal protein L7a

Por otra parte, en especies relacionadas a *T. cruzi*, como las del género *Leishmania* se ha descrito que el inicio de la transcripción ocurre en regiones de cambio de hebra o SSR por sus siglas en inglés (*switch-strand región*), especialmente en regiones SSR divergentes donde las PTUs inician en direcciones opuestas, sobre cadenas de ADN complementarias y terminan en regiones convergentes[37,184]. En el caso de *T. cruzi*, el nivel de fragmentación de los genomas ensamblados a día de hoy, así como la calidad de su anotación, complican la identificación de estas regiones. Sin embargo, los análisis de este trabajo, permitieron identificar estos sitios a lo largo de la nueva versión del genoma de la cepa SylvioX10. Se identificaron un total de 226 SSR, y hasta 15 en un solo C-LS (TcX10-chr24). Como se esperaba, todos los C-LSs contienen al menos una región divergente SSR (excepto el C-LS 17), con 5 en promedio. Existe correlación positiva (0,54) entre el número contenido de SSRs y la longitud de los C-LS, que nos permite hipotetizar sobre la uniformidad de longitud de las PTUs. Además, como se esperaba respecto al porcentaje de secuencia contenida y definida en la nueva versión del análisis con los datos de RNAseq, existe también correlación positiva entre la longitud de los C-LS, y el número de transcritos anotados (0,88) (**Figura Suplementaria 4**).

El mínimo de longitud de los transcritos (incluyendo truncados) es de 200 nucleótidos, mientras que el máximo se corresponde a un transcrito policistrónico de 20,5 kb (TCSX10.4085.2.F) que contiene 2 ORFs, y el segundo a un transcrito de 20,2 Kb que codifica una transferasa conservada de 6.635 aa. La longitud media de los transcritos está comprendida entre 800 y 2.200 nucleótidos.

#### 4.2.2 Análisis de enriquecimiento funcional del transcriptoma de *T. cruzi*

Con el objetivo de definir y poder visualizar la diversidad funcional del transcriptoma de *T. cruzi* en su conjunto a nivel molecular, se asignó una función de ontología genética (GO) a cada uno de los transcritos maduros obtenidos en este trabajo, mediante la herramienta Blast2GO y las bases de datos de InterproScan. No se pudo asignar un GO a solamente 5.048 transcritos.

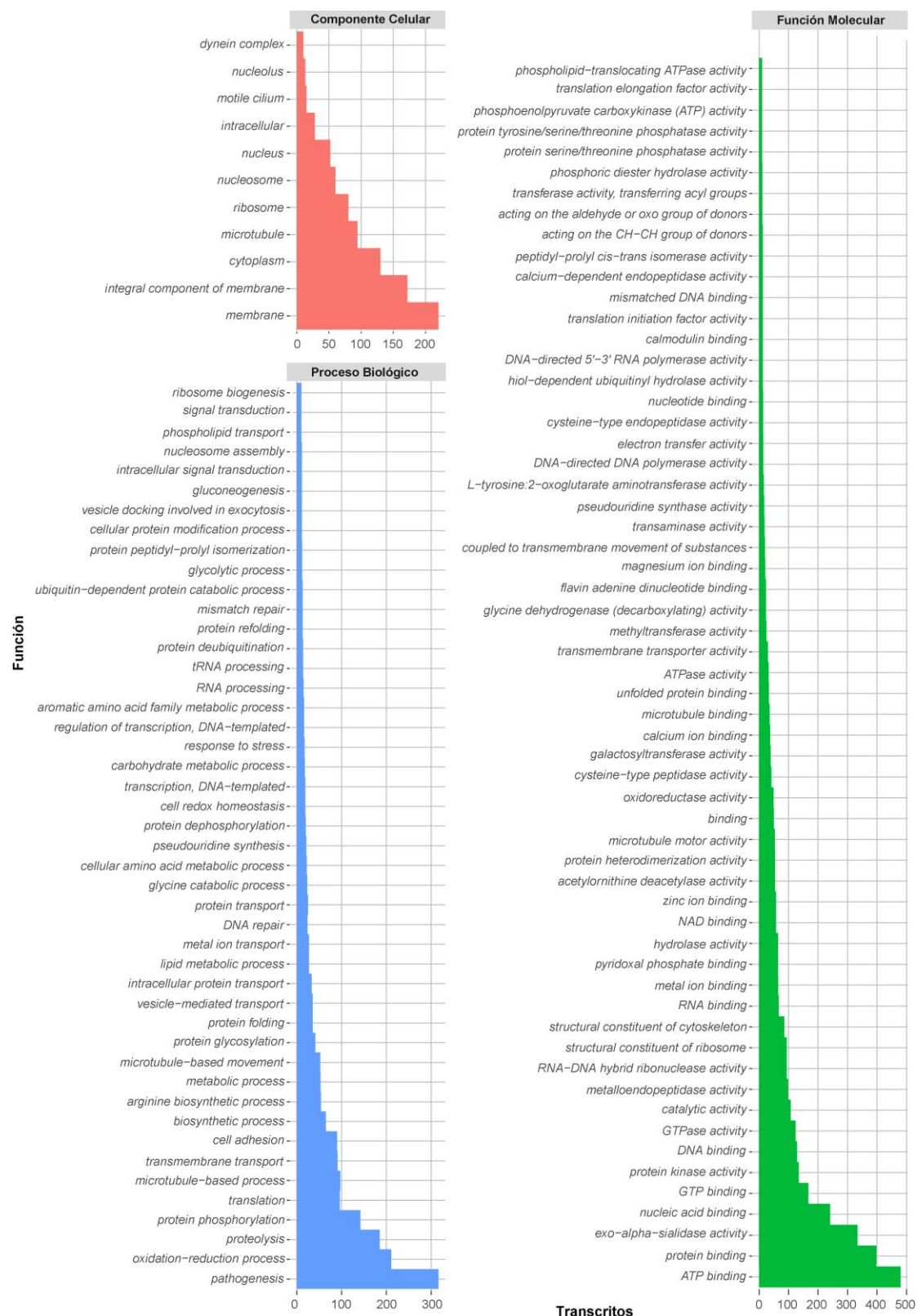
Solo los GO terms de los 3 primeros niveles y con al menos 10 transcritos fueron considerados para su visualización completa como se muestra en la **Figura 15**. De acuerdo a trabajos previos[169], se determinaron abundancias similares de transcritos ubicuos de las formas del parásito relacionados con movimiento de microtúbulos (149), organización y empaquetamiento de los cromosomas (125) y respuesta a estrés (17), entre otros.

Como se esperaba con la función “Componentes celulares”, los transcritos clasificados como productos de membrana son los más abundantes (hasta 200), seguidos de componentes del citoplasma y ribosomales clasificando un total de 1.198 transcritos en este nivel.

Los GO asociados a procesos biológicos se corresponden con objetivos específicos que el organismo está programado genéticamente a conseguir, incluyendo ARNm que están directamente o indirectamente relacionados, durante todo el proceso o en el resultado final. Se determinaron 2.812 transcritos dentro de este nivel de clasificación. Por otra parte, ya que el ARN secuenciado proviene de la forma tripomastigote (forma no replicativa, infectiva) sería de esperar que los procesos biológicos de mayor representación se correspondieran con la patogénesis (316), protección contra estrés oxidativo (210), proteólisis (185), fosforilación de proteínas (142), traducción y adhesión celular (90), entre otros procesos vitales en el ciclo de infección. Estos resultados se corresponden además con los descritos por Berná y col. (2017)[169], donde describen el enriquecimiento funcional de la forma tripomastigote (cepa no definida) establecido con procesos metodológicos similares a los empleados en este trabajo.

Finalmente, los GO del nivel “Función molecular” contienen genes asociados en actividades moleculares específicas, siendo este nivel el más enriquecido con 4.624 transcritos clasificados, dentro de los que destacan: Unión a ATP (480), unión a glicoproteínas (398), actividad sialidasa (334), proteínas de función cinasa (134) y de unión a ácidos nucleicos constituyen las 5 funciones más enriquecidas de este nivel.





**Figura 15.** Clasificación por ontología genética (primeros 3 niveles) del transcriptoma de *T. cruzi* (estado tripomastigote). Solos funciones con 10 o más transcritos fueron considerados.

Adicionalmente a este proceso de clasificación basada en ontología genética, se realizó una segunda clasificación basada en el nombre de la función contenida de cada ARNm, que permitió identificar las familias genéticas más abundantes por número de copias (**Figura suplementaria 5**).

De acuerdo con los resultados obtenidos en este trabajo con la anotación genómica de dos cepas diferentes, se determinaron un total de 1.632 familias genéticas (solo 43 más que en las cepas Bug2148 e Y), 239 de ellas con 10 o más transcritos para un total de 11.960 (aproximadamente 52% del transcriptoma total). Esta agrupación de transcritos por la función del gen codificante confirma que las funciones moleculares de *T. cruzi*, al menos en la forma tripomastigote, son cubiertas por la transcripción de genes multi-copia. La mayoría de las familias genéticas multi-copia anotadas en la secuencia genómica del parásito, son también las familias de transcritos más abundantes en la forma tripomastigote. Entre las más abundantes destacan principalmente las funciones trans-sialidasa (1.329), cinasas (338), RHS (284) y mucinas (235), entre otras.

Curiosamente, 428 transcritos corresponden a proteínas con las repeticiones *Trp-Asp* (W-D) conocidas como WD40. Estas proteínas se expresan en la mayoría de eucariotas y son implicadas en una gran variedad de funciones, principalmente transducción de señales, regulación de la transcripción, ciclo celular y apoptosis[185]. Específicamente, se ha descrito que los motivos WD40 actúan como sitios de interacción proteína-proteína, y las proteínas con estas repeticiones sirven además como plataformas para el ensamblaje de complejos proteicos o mediadoras de la interacción transitoria con otras proteínas[185,186].

Otras familias genéticas abundantes por número de copias en *T. cruzi* se corresponden con proteínas nucleósido tri-fosfato hidrolasa de la familia *P-loop* (695 transcritos). Esta familia contiene a las proteínas con el motivo nucleotídico de unión a proteínas más abundante[187], y se caracteriza principalmente por las proteínas concavalina A lectina/glucanasa que incluye además a la mayoría de glicosil hidrolasas, lectinas, liasas, beta-D-xilosidasas y vp4 de unión a ácido sialico, entre otras[188].

#### **4.2.3 Sitios de trans-splicing y poli-adenilación**

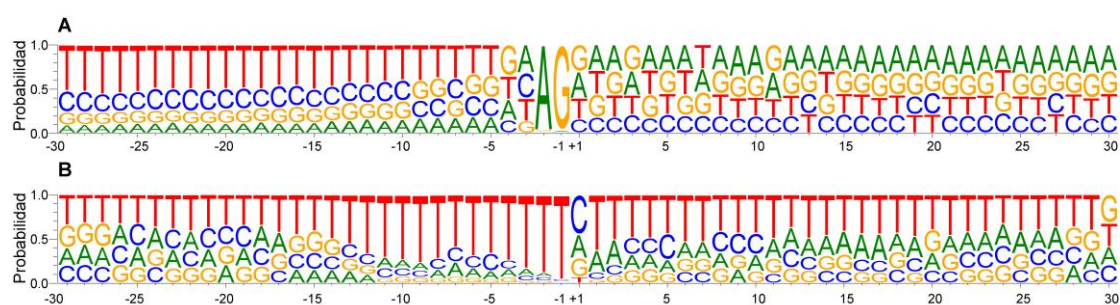
Como se mencionó anteriormente, la adición de una secuencia de 39 nucleótidos conocida como mini-exón o SL en el extremo 5' de todos los ARNm de parásitos como *T. cruzi*, constituye un proceso fundamental de maduración post-transcripcional que se le conoce como *trans-splicing*[189,190]. Actualmente se ha

descrito el dinucleótido AG en especies relacionadas a *T. cruzi*, tales como *Leishmania* y *T. brucei* como la señal nucleotídica específica necesaria para la inserción de este mini-exón y la maduración del ARN en ese extremo. Desafortunadamente, en el extremo contrario no se ha podido dilucidar una señal específica para su maduración, solamente se ha definido un enriquecimiento general de bases pirimidínicas de longitud variable[39,40,191–193].

Se han descrito pequeñas diferencias en la composición nucleotídica de las regiones próximas inmediatas al dinucleótido AG de *Leishmania* y *T. brucei*, lo que ha permitido hipotetizar, que estas variaciones a nivel de secuencia podrían generar mecanismos de control específicos en los procesos de maduración de ambas especies involucradas, causantes de enfermedades completamente diferentes[39,40]. Sin embargo, para *T. cruzi* no se había realizado ningún análisis sobre estos procesos de maduración post-transcripcionales. Por lo tanto, se analizaron las coordenadas específicas de las lecturas recortadas conteniendo las secuencias SL y poli-A, permitiéndonos además de identificar los monocistrones, analizar con resolución nucleotídica la composición de estas regiones.

Se determinó la composición nucleotídica de las regiones asociadas al SL y poli-A mediante el cálculo de la secuencia weblogo (probabilidad por frecuencia de cada base) en la región  $\pm 30$  del sitio de inserción (**Figura 16**).

Se detectaron un total de 18.589 sitios de inserción de SL, y en acuerdo con resultados anteriores en especies relacionadas[39,40], se determinó que el dinucleótido AG es también la señal más probable de *trans-splicing*, con 97,39% y 97,19% de probabilidad, respectivamente.



**Figura 16.** Composición nucleotídica (por frecuencia de base) en los sitios de inserción de SL y poli-A de *T. cruzi*. A. Sitios de inserción de SL (n = 18.589). B. Sitios de inserción de colas poli-A (n = 9.789).

Sorprendentemente, a diferencia de *L. major* (que contiene una C) pero en concordancia con *T. brucei*, un residuo de adenina es el nucleótido más probable (31,84%) en la posición -3 antes de la señal AG, seguido de una citosina (29,18%), timina (26,53%) y una guanina (12,42%). De forma igualmente interesante, en la posición -4 *T. cruzi* y *L. major* comparten una guanina como el residuo más probable con 31,2% de probabilidad, en contraste con *T. brucei* donde inicia una cadena monomérica de pirimidinas (poli-T) que continua hasta al menos otros 50 nucleótidos.

Las principales diferencias entre estas 3 especies de parásitos (*T. cruzi*, *L. major* y *T. brucei*) en las regiones próximas a la inserción de estos motivos de maduración, se basan principalmente en la composición de bases pirimidínicas. Mientras que en *L. major*, el dinucleótido C-T claramente constituye aproximadamente el 70% de la composición mayoritaria (*upstream* y *downstream*), en *T. cruzi* este patrón es diferente, conservado solamente en el extremo 5'. En el extremo 3' el patrón de secuencia es completamente opuesto, con bases púricas como composición más abundante hasta en un 60% (dinucleótido A-G) (**Figura 16A**). En el caso de *T. brucei*, se conserva el mismo patrón de composición (C-T) en la región 5' mientras que en la región 3' es diferente a las dos especies anteriores, siendo A-T la composición más probable[40]. Estos resultados indican que en general *T. cruzi* y *L. major* contienen una composición nucleotídica más proporcional en los sitios de inserción del SL (mayor complejidad lingüística) que *T. brucei*.

Contrario a lo que ocurre en la mayoría de eucariotas, los tripanosomatidos no presentan el motivo conservado AAUAA como señal de poli-adenilación[194–196]. Por el contrario, se ha descrito la presencia de dinucleótidos A-T, A-T, A-T en *T. brucei* y A-G, A-T, A-T en *L. major* en la región 3' del sitio de inserción[39,40]. Además, ha sido posible identificar una mayor variabilidad en la composición de esta región ( $\pm 40$  nucleótidos). *L. major* presenta la mayor variabilidad en la secuencia nucleotídica 5' and 3' de esta región, especialmente la 5'[39,191], mientras que en *T. brucei* es posible visualizar una composición más uniforme en ambos extremos, basada en el dinucleótido T-A como la composición más abundante.

En el caso de *T. cruzi*, se analizaron un total de 9.789 sitios de inserción de las colas poli-A (**Figura 16B**) y las regiones próximas.

Contrario a lo que se ha descrito en otros tripanosomatidos, donde no existe una señal evidente más probable como señal de la poli-adenilación, fue posible determinar un sitio específico de este proceso, marcado por un cambio brusco en la composición nucleotídica. Citosina es el nucleótido más probable con 45,32% de probabilidad, seguido de adenina y guanina en proporciones similares (26,34% y

21,52%, respectivamente) con la timina como el nucleótido menos frecuente con 6,79%.

Asimismo, este análisis permitió identificar una composición basada en timina (de hasta el 95,45% en la posición -1) y abundante en T-A (de hasta 89%) en la región 3' (posición +2).

En general, estos resultados muestran que los procesos de maduración de ARNm en *T. cruzi*, difieren de forma clara respecto a especies evolutivamente relacionadas. Al menos a nivel de secuencia y en el estado tripomastigote.

### **4.3 miARNs y transcriptómica de la fase temprana de la infección por *T. cruzi***

#### **4.3.1 Expresión diferencial de miARNs en macrófagos infectados con *T. cruzi***

En nuestro grupo de trabajo, se han establecido los tiempos más importantes en el modelo de infección *in vitro*. 0-1 hora post-infección abarca principalmente el proceso de interacción celular huésped-patógeno, a 1-6 horas incluye la internalización y finalmente de 6-24 horas después tiene lugar la proliferación intracelular del parásito[197].

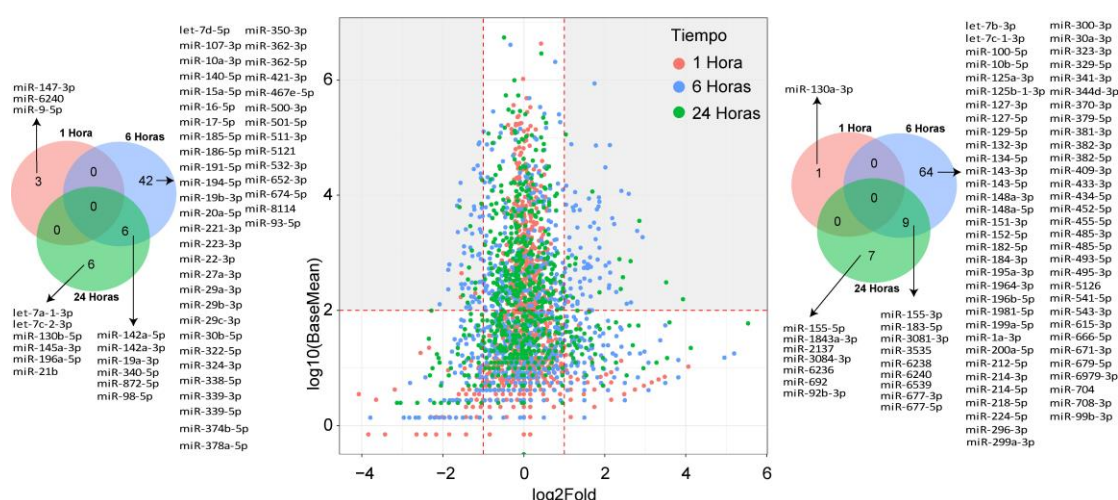
Aunque se ha descrito la participación de miARNs en respuesta a la infección por *T. cruzi*, y en general en el desarrollo de la enfermedad, actualmente no existe un estudio completo de este tipo de ARN corto en células mieloides. En este sentido se estableció como tercer objetivo de este trabajo, el análisis del perfil de expresión de miARNs en macrófagos intraperitoneales infectados con *T. cruzi* (cepa Y), en las 3 etapas fundamentales del desarrollo de la infección; interacción, internalización y proliferación.

En un primer paso exploratorio del grado de influencia de los miARNs en la infección por el parásito, se realizó la secuenciación de la fracción correspondiente a miARNs sin réplicas técnicas (3 réplicas biológicas).

Se obtuvieron en promedio 40 millones de lecturas de miARNs por muestra, con longitud comprendida entre 20 y 75 nucleótidos después del proceso de filtrado por calidad y longitud. Estas lecturas de RNAseq-dual fueron clasificadas por especie con Bowtie[116] mapeando las lecturas filtradas a ambos genomas con porcentajes de similitud superior al 95% (**Figura Suplementaria 6**). El análisis por componentes principales o PCOA por sus siglas en inglés (*Principal Components Analysis*) de las muestras de secuenciación demuestra la calidad de las muestras control, con mínima variación incluso después de 24 horas de cultivo, y la variabilidad de

expresión en las muestras infectadas con el paso del tiempo (**Figura Suplementaria 7**).

A continuación, se procesaron las lecturas correspondientes al genoma de ratón, determinando la expresión diferencial de miARNs en respuesta a la infección, respecto a las muestras control no infectadas. En la ausencia de réplicas biológicas, se usó la abundancia relativa de cada miARN como medida indicadora de su potencial expresión. Esta abundancia relativa es definida como el número de lecturas de secuenciación alineadas a cada miARN y normalizadas por la profundidad entre muestras (medida definida por DESeq2 como *BaseMean*) y expresada en décimo logaritmo (**Figura 17**). Por otra parte, la razón de cambio en los niveles de expresión de cada miARN respecto a la muestra control ( $\log_2\text{Fold}$ ), fueron calculados como la relación control/muestra de los niveles de abundancia relativa expresados en segundo logaritmo. Se estableció un *BaseMean* mínimo de 100 ( $\log_{10} = 2$ ) y un  $\log_2\text{Fold}$  mayor a 1 para considerar a un miARN como de expresión diferencial significativa.



**Figura 17.** Análisis de expresión diferencial de miARNs en macrófagos infectados con *T. cruzi*, cepa Y. Los miARNs con expresión diferencial significativa (recuadros en gris) se agrupan entre muestras con diagramas de Venn, a la izquierda con expresión diferencial negativa respecto al control no infectado y a la derecha con expresión diferencial positiva respecto al control no infectado. La razón de cambio en abundancia (lecturas normalizadas) se representa como el segundo logaritmo de la relación muestra/control.

Los resultados muestran que 6 h.p.i es el tiempo con mayor respuesta de miARNs a la infección por *T. cruzi*, con 73 miARNs sobre-expresados significativamente respecto a la muestra control, 9 de ellos además continúan con esta elevada

expresión hasta las 24 h.p.i. De la misma forma, se determinó que el mayor número de miARNs (48) con expresión menor al control ocurre a 6 h.p.i, 6 de ellos continúan inhibidos en las muestras infectadas hasta las 24 h.p.i. Las 24 h.p.i son el segundo tiempo con mayor desregulación en la expresión génica, con hasta 28 miARNs de expresión diferencial significativa (16 sobre-expresados y 12 inhibidos). Finalmente, a 1 h.p.i se determinó el menor cambio en el perfil de expresión con solo 4 miARNs deregulados (solo uno con sobre-expresión).

Los niveles de miR-130a-3p ( $\log_2\text{fold} = 1,09$ ) a 1 h.p.i contrastan con los niveles del miR-142-5p ( $\log_2\text{fold} = -0,06$ ). En el trabajo publicado por Shicheng Su y col.[198] se describió que este desbalance puede causar la inhibición de IL-4 e IL-13, afectando la capacidad pro-fibrogénica en macrófagos. Sin embargo, se ha descrito que las dianas de un miARN dependen del contexto celular y estos resultados podrían no ser comparables, principalmente debido a diferencias biológicas (macrófagos de médula ósea, y el efecto de la infección) con ese trabajo. Por lo tanto, se necesitan más análisis para confirmar una posible hipótesis pro-fibrogénica inducida por la infección.

Para determinar el potencial efecto de los miARNs en la expresión génica a 6 h.p.i, se realizó un análisis de enriquecimiento funcional de los transcritos diana de los miARN, calculados con Diana-mirPath v3[119] (**Figura 18**). Se consideraron solamente los miARNs significativos (sobre-expresados [**Figura 18A**] o inhibidos [**Figura 18B**] por separado).

De forma interesante, estos resultados sugieren que algunas de las rutas implicadas en el desarrollo de la infección y de la enfermedad, podrían ser regulados por la expresión de los miARNs. Estadísticamente, la ruta más afectada por la sobreexpresión de miARNs (10) es “*gap junctions*” (GJs) con 14 genes potencialmente afectados, entre ellos *Gja1* (*Gap Junction alfa 1*, también conocida como *Cx43*) y *Gjd2* (*Gap junction delta 2*) (**Figura 18A**). Los genes asociados a esta ruta tienen como principal función la comunicación intercelular, pero en corazón desempeñan un papel fundamental en la homeostasis metabólica y sincronización rítmica de las contracciones de los miocitos cardiacos[199].

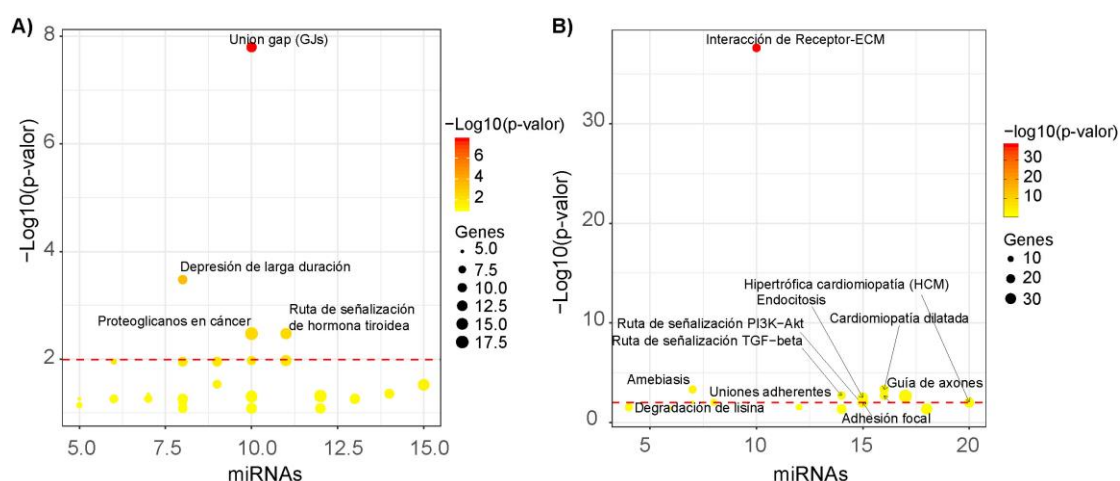
En este sentido, se ha demostrado que *T. cruzi* es capaz de causar alteraciones en este mecanismo de comunicación célula-célula y que la disminución en la expresión de la *Gja1* está relacionada con la proliferación del parásito y daño en tejido cardiaco[199]. Sin embargo, la desregulación de la expresión génica por miARNs sugiere que la alteración de este mecanismo podría incluir a un grupo genético más amplio (Resultados completos; **Tabla Suplementaria 1**).



Igualmente, se ha demostrado que la producción de citoquinas de respuesta inmune por el huésped a la infección por *T. cruzi* puede causar trastornos depresivos[200]. Sin embargo, los genes específicos de esta afectación no han sido identificados. En este trabajo, se identificaron 8 miARNs con 10 dianas diferentes que podrían estar relacionados con la afectación de la vía “depresión prolongada” (*long-term depression*) (Resultados completos; **Tabla Suplementaria 1**).

Con 11 miARNs y 14 ARNm diana, la señalización por hormonas tiroideas es otra de las vías potencialmente afectadas por la desregulación de miARNs. Casi la mitad (*Med1*, *Med13*, *Med14*, *Med121* y *Med27*) de estas dianas pertenecen al complejo mediador transcripcional y receptores de hormonas tiroideas, lo que concuerda con el trabajo publicado por Bellera y col. (2014)[201] donde se describe que el tratamiento con levotiroxina de pacientes Chagásicos (aumenta los niveles de hormonas tiroideas) es capaz de disminuir los niveles de proliferación intra-celular del parásito.

Finalmente, pero no menos relevante la 4 ruta metabólica afectada por la sobreexpresión de miARNs es “proteoglicanos en cancer”, con 18 dianas y 10 miARNs (Resultados completos; **Tabla Suplementaria 1**). En este caso, la mayoría de genes se encuentran involucrados en la conservación, generación y/o mantenimiento de la matriz extracelular del huésped, una de las principales barreras para la interacción e internalización de patógenos, pero que parásito es capaz de modular en su beneficio[202] (**Figura 18A**).



**Figura 18.** Análisis de enriquecimiento funcional de los ARNm regulados por la expresión de miARNs a 6 h.p.i. **A.** Enriquecimiento funcional de los ARNm inhibidos por los miARNs sobre-expresados significativamente. **B.** Enriquecimiento funcional de los ARNm expresados en ausencia de su miARN regulador. P-valor menor o igual a 0.01.



La regulación y uso de la matriz extracelular (MEC) del huésped mamífero por el parásito, ha sido uno de los temas centrales de investigación en la fase temprana de la infección[202]. Se ha demostrado que las glicoproteínas *Gp83*, *Tc85*, *P45* presentes en la membrana del parásito son capaces de activar la expresión de la laminina  $\gamma$ -1 (*Lamc1*) y trombospondina-1 (*Thsb1*) facilitando su interacción y capacidad de internalización. Estudios *in vitro* muestran que la inhibición de la *Lamc1* y *Thsb1* usando ARNi reducen significativamente esta capacidad[203,204]. Coincidentemente con estos resultados previos, la ruta más enriquecida por la inhibición de miARNs y la consecuente expresión de los ARNm diana es la relacionada a la MEC (**Figura 18B**). Los resultados de este trabajo muestran que la activación en la expresión de los genes relacionados con las rutas de *Lamc1* (*Sv2c* e *Itgb8*) y *Thsb1* (*Itgav* y hasta 11 proteínas de colágeno diferentes) que facilitan la internalización del parásito, podrían estar regulados por la inhibición de los miARNs de la familia miR-29 (a, b y c de la región 3p), principalmente (Resultados completos; **Tabla Suplementaria 1**). Estos datos confirman resultados previos, pero amplían el repertorio genético potencialmente involucrado en la activación de la ruta por la inhibición en la expresión de miARNs, y sugieren el rol dual de la MEC durante la infección, por una parte, el parásito es capaz de modelarla (degradando varios de sus componentes) promoviendo una mayor interacción con la membrana del huésped y por otra, como resultado de esa interacción activar la expresión de genes del huésped en su beneficio.

Curiosamente, las rutas significativamente activadas por la inhibición en la expresión de miARNs a 6 h.p.i, se corresponden a las rutas más ampliamente estudiadas en la fase temprana de la infección (principalmente *PI3K signalling* [PI3K], *TGF- $\beta$  signalling*, *endocytosis*, *adherens junction* [ADJ], *focal adhesión* y *lysine degradation*) y otras estrechamente relacionadas (*Amoebiasis*, *Axon guidance*, *Dilated cardiomyopathy* [DCM] y *Hypertrophic cardiomyopathy* [HCM]) con la participación de los mismos genes[202,205–209] (Resultados completos; **Tabla Suplementaria 2**). En general, los resultados sugieren que la inactivación de 29 miARNs en la fase temprana de la infección y la consecuente expresión de sus genes dianas (111 únicos), puede causar la activación de estas rutas involucradas en procesos de señalización, reorganización del citoesqueleto, respuesta inmune e interacción huésped patógeno, que en su mayoría convergen en una mayor capacidad de internalización y mayor proliferación intracelular por el parásito.

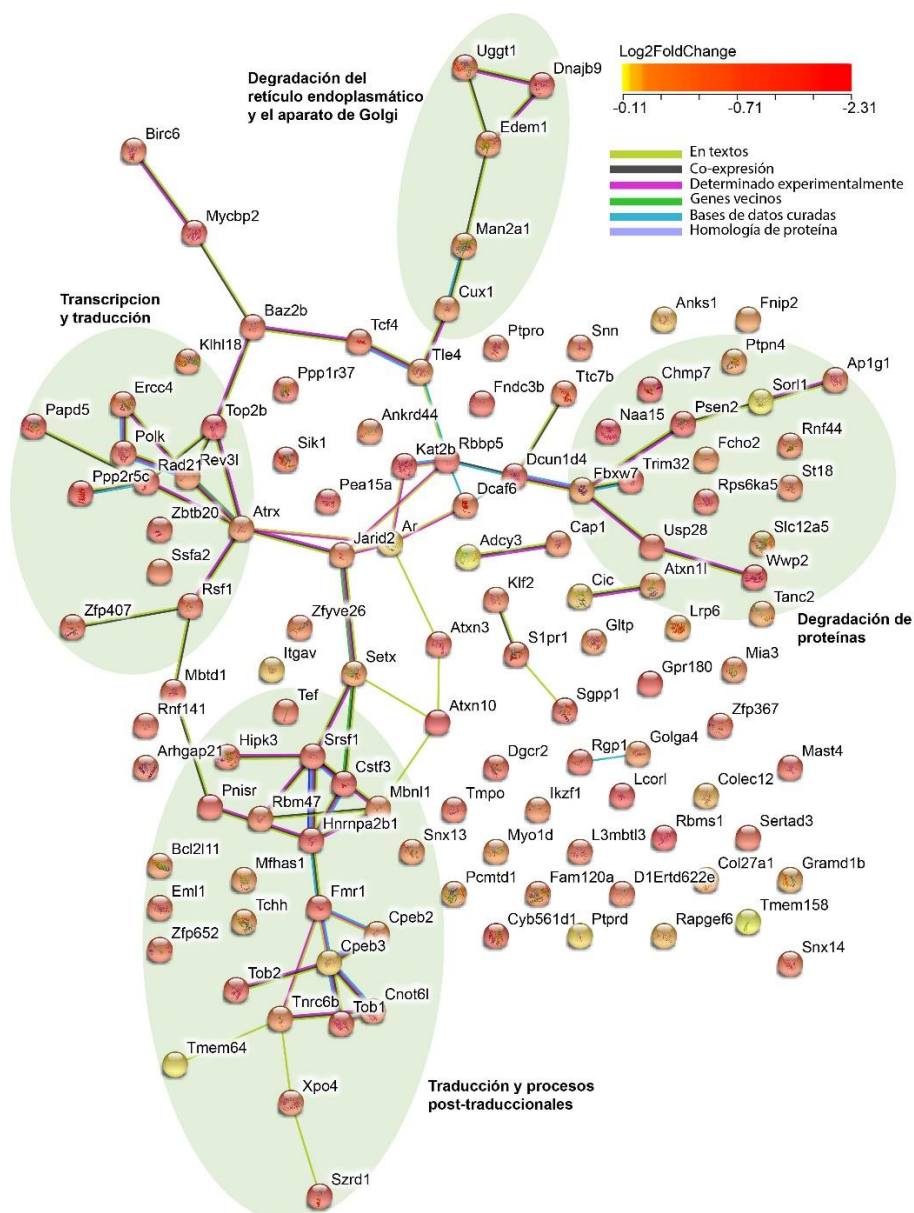
La ruta PI3K resulta enriquecida por la disminución en la expresión de 17 miARNs y el consecuente aumento en expresión de al menos 37 genes (15 específicos de la ruta PI3K), entre ellos la subunidad reguladora *Pik3r3* de la proteína Pi3K. Como era

de esperar, la ruta de señalización de TGF- $\beta$  es activada por aumentos en la expresión de *Tgfb1* y *Tgfb2* y otros 12 genes involucrados en esta ruta, coincidentemente con resultados previos 3 miARNs (miR-322-5p, miR-338-5p, miR-223-3p) son específicos de la activación de la ruta, mientras que los genes y miARNs restantes están involucrados en la activación de rutas complementarias/relacionadas con el proceso de interacción/internalización del parásito y desarrollo de la enfermedad, tales como la vía de axones, endocitosis, adhesión focal, ADJ, PI3K, HCM y DCM[207,210].

De forma general, el enriquecimiento funcional de rutas estrechamente relacionadas a 6 h.p.i por la disminución en la expresión de diversos miARNs, es resultado de que en su mayoría estas rutas comparten los mismos genes. Solamente TGF- $\beta$  (miR-322-5p, miR-338-5p, miR-223-3p), endocitosis (miR-194-5p, miR-22-3p) y la vía de axones (miR-511-3p, mmu-miR-339-5p) podrían estar regulados negativamente por miARNs específicos (**Figura Suplementaria 8**). En este sentido, cabe señalar que el enriquecimiento funcional de rutas relacionadas con procesos propios del desarrollo de la enfermedad más que con la fase inicial de la infección (fase aguda y crónica; HCM, DCM, principalmente) se debe a esta redundancia de miARNs y dianas ya que no es posible explicar alteraciones cardíacas en este estudio realizado con macrófagos.

Por otra parte, de los 16 miARNs sobre-expresados a 24 h.p.i solo 6 son específicos de esta etapa, los 9 restantes se comparten con los miARNs significativos a 6 h.p.i. Las dianas de estos miARNs compartidos (principalmente miR-677-3p, miR-677-5p, miR-6539 y miR-183-5p), se relacionan con el enriquecimiento funcional de las rutas depresión prolongada y GJs, como era de esperar debido a la activación de la respuesta inmune y el daño acumulado en la MEC.

Por la baja cantidad de genes inhibidos con el aumento de expresión de miARNs a 24 h.p.i (121), no es posible realizar un análisis de enriquecimiento funcional significativo. Sin embargo, se realizó una búsqueda de las potenciales relaciones biológicas entre estos ARNs (**Figura 19**).



**Figura 19.** Interacciones conocidas y predichas (físicas y funcionales) entre los genes diana de miARNs sobre-expresados a 24 h.p.i. Se remarcan los grupos genéticos funcionalmente relacionados.

Estos resultados sugieren que las rutas y/o procesos afectados con el aumento en la expresión de miARNs a 24 h.p.i son aquellos involucrados en procesos transcripcionales, traduccionales y post-traduccionales, pero, además es posible que la degradación del retículo endoplasmático y del aparato de Golgi, un proceso recientemente relacionado con la formación de la PV (*Parasitophorous vacuole*)[211] este regulado por la sobre-expresión de miARNs y la disminución de sus dianas (*Uggt1*, *Dnajb9*, *Edem1*, *Man2a1* y *Cux1*).

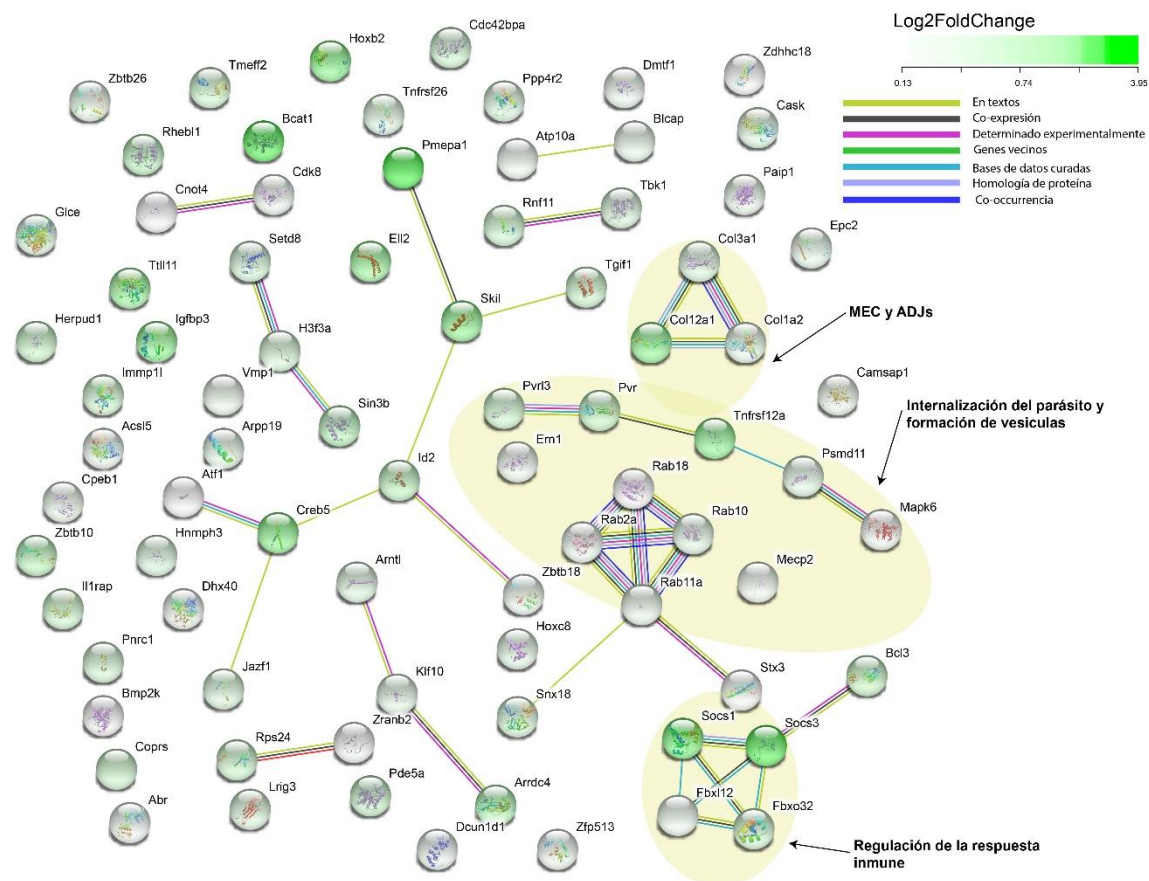
Finalmente, 6 de los 12 miARNs con menor expresión respecto al control no infectado a 24 h.p.i, se comparten a 6 h.p.i (miR-142a-3p, miR-142a-5p, miR-19a-3p, miR-340-5p, miR-872-5p, miR-98-5p). Como es de esperar, estos miARNs compartidos están relacionados con la regulación negativa de diversas vías de internalización y respuesta inmune, principalmente Tgf- $\beta$ , PI3K y endocitosis por la activación de sus dianas *Col1a2*, *Col3a1*, *Rab11a*, *Id2* y *Creb5*.

De la misma forma que con los miARNs sobre-expresados a 24 h.p.i, las dianas sobre-expresadas por la ausencia de su miARN regulador a 24 h.p.i no son suficientes para calcular un enriquecimiento funcional significativo, sin embargo, el análisis de sus interacciones (usando STRING[120]) permitió dilucidar las potenciales vías afectadas (**Figura 20**).

Como se mencionó anteriormente, 3 tipos de colágenos involucrados en el proceso de internalización del parásito mediante la alteración de las vías de mantenimiento y estructura de la MEC, adhesión focal, ADJs y axones están reguladas por los miARNs miR-98-5p, miR-29b-3p. La activación de los genes *Rab18*, *Rab10*, *Rab11a*, *Rab2a*, *Tnfrsf12a* y *Pvr13*, a su vez, permite hipotetizar sobre la continuidad del proceso de internalización del parásito mediante su fagocitosis y la activación de la ruta PI3K como se describió anteriormente por Maganto-García y col. (2008)[212] con *Rab5*. Aunque, los resultados del presente trabajo no presentan evidencia de la activación de la *Rab5*, se ha demostrado que los genes de la familia Rab regulan el tráfico de membrana incluyendo la formación de vesículas que podrían favorecer la internalización del parásito[213].

Hovsepián y col. (2013)[214] demostraron que *T. cruzi* es capaz de modular la respuesta inmune del huésped (en cardiomiocitos) mediante el aumento de la IL-10 que promueve la fosforilación de STAT3, aumenta los niveles de SOCS-3 (*Suppressor of Cytokine Signalling-3*) evitando la translocación de NF- $\kappa$ B y la fosforilación de ERK1/2. Sin embargo, nuestros resultados sugieren que esta capacidad de regulación de la respuesta inmune podría extenderse a SOCS-1 (**Figura 20**) mediante la activación de STAT1, un regulador negativo de IFN $\gamma$  indispensable en la eliminación del parásito intracelular[215].

Cabe destacar la presencia de miARNs capaces de controlar la expresión de genes implicados en hipertrófia cardíaca y cardiopatía dilatada, enfermedades muy similares a la cardiopatía chagásica.



**Figura 20.** Interacciones conocidas y predichas (físicas y funcionales) entre los genes diana de miARNs inhibidos a 24 h.p.i. Se remarcen los principales grupos genéticos funcionalmente relacionados.

El análisis individualizado de algunos miARNs, muestra niveles alterados de varios miARNs implicados en la diferenciación M1/M2 de macrófagos. Por ejemplo, niveles elevados de let-7c-1-3p y miR-143-3p y bajos de miR-9-5, miR-27a o miR-19a-3 se relacionan en la diferenciación M2 de macrófagos y en la respuesta microbicida contra la infección de *Mycobacterium*[216,217]. Algunos otros como miR-127-5p promueven la inhibición de *Irf8* y la inducción de *Arg1* e *iNOS* en las MDSC[218]. miR-155-3p y miR-100-5p que se encuentran elevados promueven la inhibición de *Tlr4*[219–222], un receptor necesario para la eliminación de *T. cruzi*. Macrófagos que sobreexpresan miR-148a-3p, como los infectados a 6 horas en este trabajo, presentan una mayor capacidad fagocítica y expresan mayores cantidades de ROS[223]. La sobreexpresión de miR-455-5p está implicado en la diferenciación M2 y la inhibición de *Socs3*[224]. En su conjunto, estos resultados sugieren que la infección por *T. cruzi* es capaz de evitar una respuesta M1, necesaria para su eliminación[44].

### 4.3.2 Análisis de la expresión génica en macrófagos infectados a 24 h.p.i

Con el objetivo de validar los resultados obtenidos por miARN-seq, se realizó un estudio de la transcriptómica (ARNm, por triplicado) del huésped en respuesta a la infección por *T. cruzi* a 24 h.p.i.

Después del filtrado por calidad, longitud y eliminación de adaptadores de secuenciación, se obtuvieron un total de 45 millones (en promedio) de lecturas pareadas y de 75 nucleótidos de longitud por muestra. Las muestras control resultaron con un porcentaje de alineamiento entorno al 94% al genoma del huésped, mientras que las muestras infectadas un 6% menos (88% en promedio). Las muestras de cada grupo (control e infectado) mostraron un alto grado de reproducibilidad y claras diferencias en sus perfiles de expresión (**Figura Suplementaria 9**).

Mediante el análisis de transcriptómica fue posible identificar la expresión diferencial significativa ( $FDR \leq 0,001$ ) de 5.292 genes, de los cuales 115 muestran una sobre-expresión mayor o igual a 8 órdenes de magnitud respecto a la muestra no infectada ( $\log_2\text{Fold} \geq 3$ ) y 36 con expresión disminuida (**Figura Suplementaria 10**).

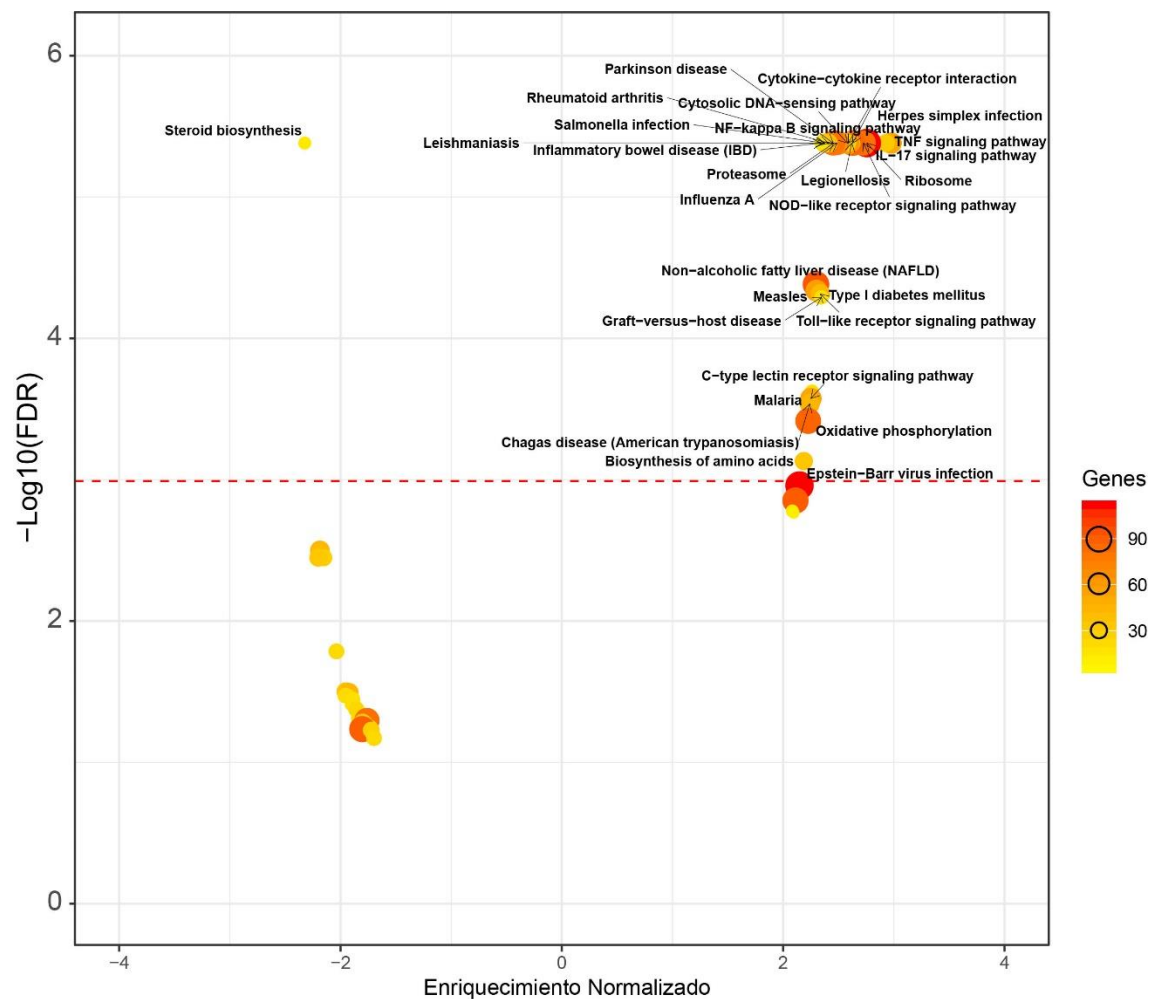
En general, en la **Figura Suplementaria 10** que resume el resultado de la expresión diferencial, es posible identificar a algunos de los genes más ampliamente estudiados en la infección por *T. cruzi* principalmente en la fase temprana, tales como *Tnf*, *Il6*, *Il1b*, *Acod1*, *Cxcl3*, *Nfkb*, *Tlr2*, *Irf7*, *Cxcl6*, *Arg2*, *Tgfb*, *Socs3*, entre otros. Sin embargo, para determinar en su totalidad las rutas biológicas potencialmente involucradas y/o activadas en esta fase de la infección, se realizó el análisis de enriquecimiento funcional con los genes significativamente expresados (5.292) (**Figura 21**).

Tal y como se describió anteriormente, en un análisis de enriquecimiento funcional los resultados están condicionados a las rutas y genes más analizados, así como a la redundancia generada por la participación (directa o indirecta) de un gen en más de una ruta. Por ejemplo, en nuestro análisis de los 62 genes con sobre-expresión asociados con la enfermedad de Parkinson, 50 están relacionados también con procesos de la fosforilación oxidativa, la infección por el virus Epstein-Barr y NAFLD (*Non-Alcoholic Fatty Liver Disease*), y para esta última (NAFLD) solo 5 de sus 52 genes identificados no han sido asociados significativamente a otros procesos. De esta forma, es posible omitir justificadamente la atención a algunas de las rutas significativamente enriquecidas, pero sin relación biológica evidente con el modelo

de infección *in vitro* empleado en este trabajo tales como la enfermedad de Parkinson, artritis reumatoide, diabetes mellitus tipo 1, sarampión y NAFLD.

De esta manera, las rutas funcionalmente enriquecidas en las muestras infectadas se corresponden en su mayoría con rutas descritas anteriormente, asociadas al proceso de infección y/o desarrollo de la enfermedad de Chagas, así como a algunos otros tipos de infección y/o enfermedades relacionadas en los que se desarrollan procesos similares de interacción huésped-patógeno, internalización y activación de la respuesta inmune (Leishmaniasis, influenza, infección por salmonella, herpes, legionelosis y malaria, principalmente). Por otra parte, con 12 genes inhibidos y un enriquecimiento negativo normalizado de -2.32, la biosíntesis de esteroides es la única ruta significativamente inhibida en este modelo de infección (**Tabla Suplementaria 3.**).

Este análisis transcriptómico, nos permitió confirmar la inhibición y sobre-expresión de genes en respuesta a los niveles de miARNs descritos en el capítulo anterior, con un porcentaje de falsos positivos menor al 1%, confirmando que el uso del algoritmo de Diana tools[225] (microT-CDS) y su base de datos de interacciones (TarBase v8) es un método fiable para la predicción de ARNm dianas.



**Figura 21.** Análisis de enriquecimiento funcional de los genes expresados diferencialmente a 24 h.p.i. Macrófagos intraperitoneales de ratón BALB/c infectados con *T. cruzi*.

Los resultados de transcriptómica demuestran que a 24 h.p.i el 1,53% (81) de los transcritos con sobre-expresión significativa ( $p$ -valor ajustado  $\leq 0,001$ ) están regulados por la inhibición de 12 miARNs, mientras que el 2,53% (121) de los transcritos significativamente inhibidos ( $p$ -valor ajustado  $\leq 0,001$ ) dependen de la sobre-expresión de al menos 16 miARNs. Y aunque no es posible descartar el importante rol de los miARNs en el desarrollo de la enfermedad de Chagas a este tiempo post-infección (principalmente por la regulación negativa de genes como *Bcat*, *Tnfrsf12a*, *Bcl3*, *Socs3*, *Psmc11*, entre otros), los resultados muestran que la mayor regulación en la expresión génica de los macrófagos infectados por miARNs ocurre a 6 h.p.i.

Por otra parte, el análisis de la transcriptómica a 24 h.p.i mediante el enriquecimiento funcional permitió ampliar de manera significativa una ruta completa de activación de



la respuesta inmune en la infección (**Figura 22**). De forma general, los genes implicados en la activación de la respuesta inmunitaria son un resultado de la intersección de las rutas funcionalmente enriquecidas, principalmente *Tnf*, *Nfkb*, *Nod*, *Tlr*, y la propia propuesta para la enfermedad de Chagas por la enciclopedia de genes y genomas Kyoto ó KEGG por sus siglas en inglés (*Kyoto Encyclopedia of Genes and Genomes*, ruta map05142).

En acuerdo con resultados anteriores, el análisis de transcriptómica demuestra que la respuesta inmune es dependiente en gran parte del reconocimiento de diversos PAMPs en la superficie, principalmente mediante *Tlr2* y de *Tgf-βr*, pero por primera vez en *T. cruzi*, los resultados sugieren la expresión de *Cd14* como co-receptor de *Tlr2* en la ausencia de *Tlr4*, *Tlr6*, *Tlr1* y la sobre-expresión del gen *Lbp* por el reconocimiento de lipoproteínas del parásito. Además, se determinó una amplia ruta de activación dependiente de *Myd88*, que incluye la expresión de *Irak3* (antes solo 1 y 4) y 5 genes de la familia *Traf* (antes solo *Traf6*) y la consecuente activación de 5 map-cinasas adicionales (antes solo 4) que promueven la expresión de los factores de transcripción *Fos* y *Jun* activando la producción de citoquinas inflamatorias y pro-inflamatorias.

Silva y col. (2010)[226], describieron que la infección por *T. cruzi* en modelos *in vitro* *Myd88*–/– producen los mismos niveles de citocinas que los modelos WT, sugiriendo la existencia de una vía alternativa de respuesta inmune independiente de *Myd88* y potencialmente dependiente de *Nod1* mediante el reconocimiento de ADN del patógeno. Los resultados del presente trabajo sugieren que esta vía independiente puede ser activada por la expresión de *Ticam2* dependiente de *Tlr2* y hasta 4 genes de la familia *Traf* pero, además, por la activación de *Tlr7* y *Nod2* en respuesta a la detección de ARN del patógeno, confirmando suposiciones previas sobre la importancia de *Tlr7* en la respuesta inmune[227]. Cabe resaltar que, aunque se ha descrito anteriormente la expresión de *Tlr9* y *Nod1* por el reconocimiento de ADN exógeno en la fase temprana de la infección, no fue posible detectar su expresión a pesar del evidente aumento en la carga parasitaria (por aumento en los niveles de ARNm del parásito), indicando una sorpresiva importancia del material genético del parásito (ARN) en el desarrollo de la respuesta inmune, aparentemente independiente de su proliferación.

Por otra parte, con base en los resultados, se sugiere por primera vez en *T. cruzi* la participación de *Irf9* en la fase temprana de la infección como potencial activador de interferones de tipo 1 junto con *Irf7* e *Irf3* contemplados anteriormente[44]. Es necesario señalar que la función de estos genes se extiende a otras rutas, tales como la activación de *Nfkb* la vía *Stat1*.



significativa respecto a la muestra control y que nuestro grupo de trabajo ha determinado que juega un papel muy importante en la infección de macrófagos por *T. cruzi*[197]. *Irg1* junto a *Bcat1* son 2 de las proteínas más importantes en la reprogramación metabólica de macrófagos y que claramente afectan sus propiedades inflamatorias/anti-inflamatorias[228,229]. C) Quimiocinas: *Cxcl3*, *Cxcl2*, *Ccl5*, *Cxcl1*, *S100a8*, *Cxcl16* y *Rarres2*. D) Receptores: *Clec4e*, *Siglece*, *Itgal*, *Pilrb1*, *Pilrb* y *Pilra*, D) Metaloproteasas *Spata13*, *Col5a3*, *Timp1* y *Mmp2*. E) Relacionados a TGFβ *Rnd1*, *Tmeff1*, *Pmepa1*, *Acp5* y *Tgfb1*.

Algunos de los genes más alterados en nuestro análisis (*Irg1*, *Nfkb1a*, *Nfkbiz*, *IL-1b*, *Socs3* y *Ccl5*) se han encontrado inducidos en células de microglia durante la infección por *Listeria*[230], que presenta ciertas similitudes a la infección por *T. cruzi*. Aunque es bien sabido que la respuesta inmune a *T. cruzi* en modelos *in vitro* es un proceso multifactorial, dependiente principalmente de la línea celular y la cepa del parásito, nuestros resultados constituyen un avance significativo en el establecimiento del perfil genético resultante de la interacción huésped-patógeno y la proliferación del parásito. Por una parte, se han confirmado gran parte de los genes encargados de producir la respuesta inmune y por otra, se amplía de forma considerable su relación y/o dependencia biológica.

#### 4.3.3 Perfil transcriptómico del parásito 24 h.p.i

Finalmente, aunque menos del 6% de los datos de secuenciación en las muestras infectadas se corresponde con el material genético del parásito, su separación *in silico* del total, permitió determinar su perfil transcriptómico más abundante (**Figura Suplementaria 11**). Aunque en nuestro modelo de infección el transcriptoma del huésped es mucho más abundante y por tanto causa el enmascaramiento del transcriptoma del parásito, los resultados nos permiten visualizar de manera general sus transcritos más abundantes en el desarrollo de la infección. De esta manera, los resultados se corresponden con lo esperado y con los resultados del transcriptoma de la forma infectiva de Sylvio X10 de este trabajo. Los transcritos más abundantes se corresponden a genes ribosomales, histonas, chaperonas, choque térmico, transporte entre otras asociadas a los procesos de transcripción, traducción y metabolismo. Además, se identificaron 8 ARNms correspondientes a mucinas (principal ligando de *Tlr2*) y uno con actividad trans-sialidasa. De forma igualmente interesante, fue posible identificar la expresión de la Calreticulina, una proteína implicada en la internalización y evasión de la respuesta inmune[144].

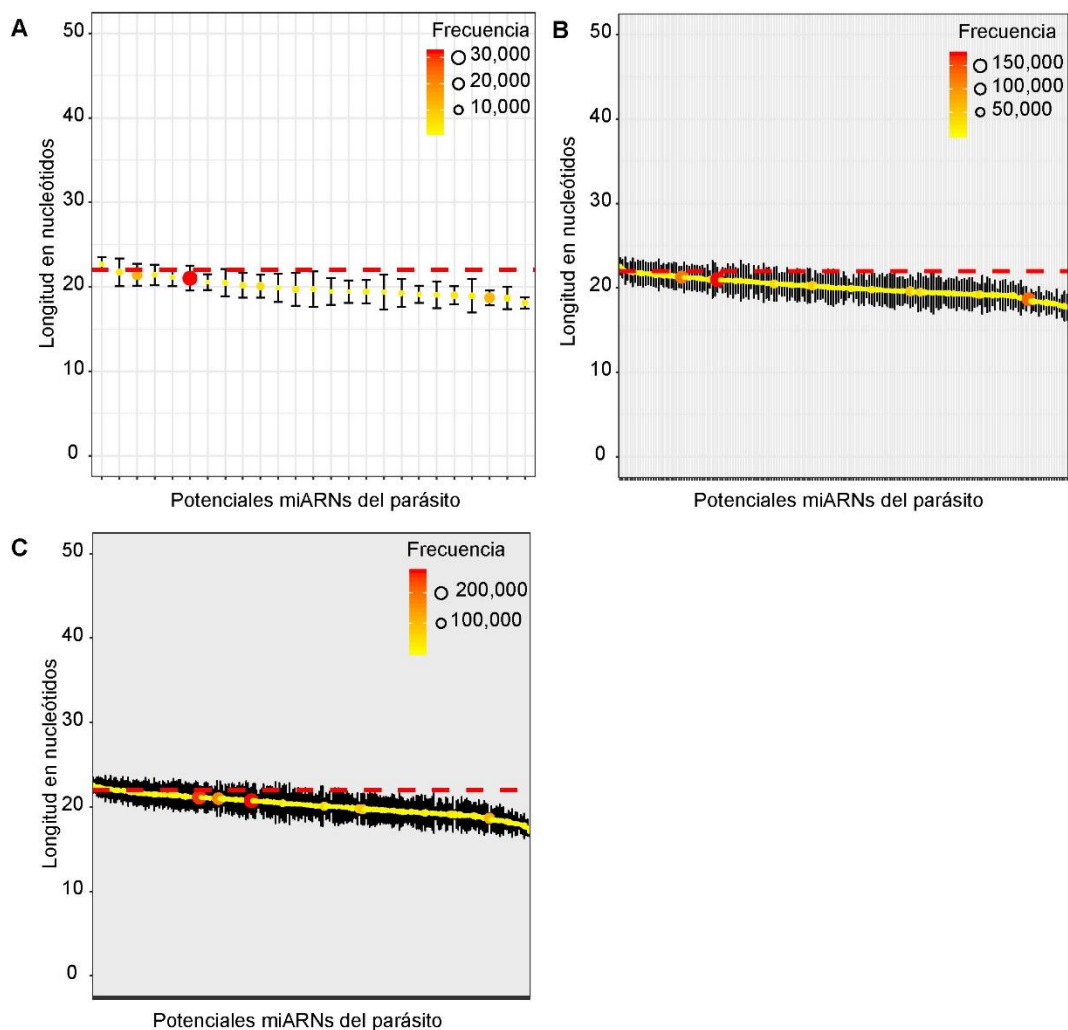
#### 4.3.4 Presencia de ARNs cortos de *T. cruzi* con potencial función de miARNs

En la sección anterior se hizo referencia a un proceso de enmascaramiento en el ARNseq dual (huésped-parásito), causando que menos del 6% de los resultados de secuenciación se corresponda al material genético del patógeno. Este proceso ocurre debido a la diferencia en abundancia de los tipos celulares, aunque ambas especies contienen un transcriptoma total aproximadamente equivalente (entorno a los 20 mil transcritos), la cantidad de macrófagos sigue siendo mayor que la de parásitos después de 24 horas. Es un resultado que se corresponde con lo esperado. Sin embargo, en el análisis de miARNs este patrón de abundancia es completamente diferente, después de 24 horas de la infección más de 7 millones de lecturas (36,6% del total) se corresponden al material genético de *T. cruzi* (**Figura Suplementaria 6**). Este aumento sustancial en la cantidad de ARN pequeño del huésped es de especial interés por dos razones, primero porque es independiente de la cantidad de parásitos (32,9 % a 6 h.p.i) y segundo porque la presencia de ARN del patógeno puede ser capaz de modular la respuesta inmune del huésped, mediante *Tlr7* y *Nod2* como se describió en la sección anterior. Por lo tanto, las lecturas de secuenciación a 1, 6 y 24 horas (miARNseq) correspondientes al genoma del parásito fueron separadas *in silico* y analizadas investigando sus potenciales implicaciones biológicas.

Sorprendentemente, el perfil de la distribución de longitud de las lecturas correspondientes al genoma de *T. cruzi* en el miARNseq, presenta exactamente el mismo patrón que para el huésped (**Figura Suplementaria 12**), aun cuando se ha descrito que la mayoría de parásitos (especialmente kinetoplastidos) no tienen la maquinaria molecular necesaria para su biogénesis y uso (genes *Dicer* y del tipo *Argonauta*)[68,75], sugiriendo en primera instancia que, si el parásito en realidad es capaz de producir miARNs debería hacerlo a través de un mecanismo de biogénesis alternativo desconocido hasta el momento, o por la propia intervención de la maquinaria molecular del huésped. Por otra parte, también existe la posibilidad de que este ARN del parásito sea producto de su muerte y la consecuente degradación de su material genético, y su existencia sea debida a las condiciones de la preparación de las librerías que no pueden excluirlo, así como la alta estabilidad de ARNs cortos. Para eliminar esta posibilidad se determinó si el ARN con longitud teórica de miARNs maduros (15-25 nucleótidos) contenían secuencias sobre-representadas, como una forma indirecta de calcular los niveles de abundancia de cada secuencia y descartar su formación como producto del azar por degradación (**Figura 23**).

Las lecturas de secuenciación correspondientes al genoma del parásito fueron agrupadas por similitud de secuencia (superior al 90%) y descartando grupos con menos de 1.000 lecturas. Los resultados demuestran que existen potenciales miARNs del parásito con cobertura de secuenciación superior (cobertura de hasta 953 mil lecturas a 24 h.p.i y de hasta 685 mil a 6 h.p.i) a algunos miARNs constitutivos del huésped (mir-23a-3p: aproximadamente 200 mil secuencias en las muestras control), principalmente a 6 y 24 horas post-infección (**Figura 23. B, C**). Se identificaron un total de 175 grupos o Tc-mirs (potenciales miARNs de *T. cruzi*) a 24 h.p.i, 113 a 6 h.p.i y 21 a una 1 h.p.i.

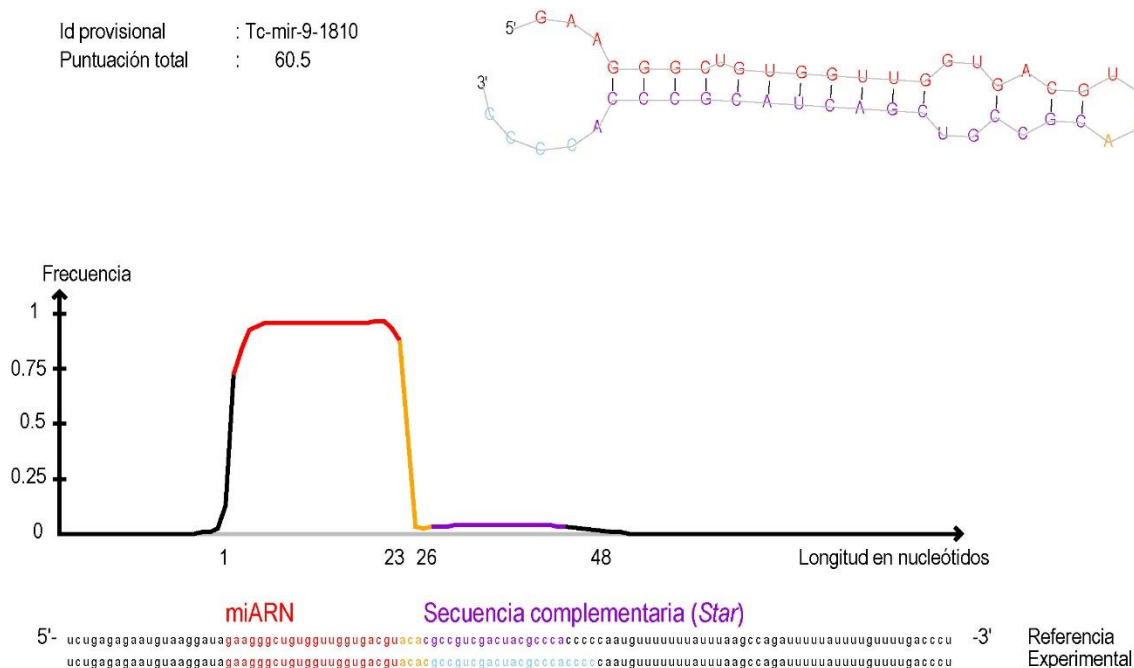
Sin embargo, para considerar que fragmento de ARN cualquiera puede generar un miARN maduro y funcional por una ruta canónica o alternativa, existen una serie de consideraciones adicionales y rigurosas inherentes a los procesos de biogénesis conocidos que deben cumplirse satisfactoriamente. MiRDeep2[121] es un software bioinformático diseñado específicamente para el análisis de datos de miARNseq, que incluye la detección de potenciales nuevos miARNs, siendo probablemente al día de hoy el software más empleado a nivel mundial para este propósito. Las consideraciones más importantes de este software son la detección de la secuencia completa del miARN-precursor no solo del miARN maduro, el cálculo de su estructura secundaria, la conservación de la secuencia semilla (*seed-sequence*) con otras especies (la mayoría de miARNs conocidos son altamente conservados entre especies) y su capacidad de auto-estabilizarse formando la estructura *loop* característica de los miARN-precursores. Estas características son fundamentales para la interacción con la maquinaria de procesamiento y el RISC (principalmente *Dicer* y *Ago2*).



**Figura 23.** Lecturas de miARNseq dual correspondientes al genoma del parásito con longitud de miARNs agrupadas por homología de secuencia superior al 90%. A. 1 hora post-infección. B. 6 horas post-infección C. 24 horas post-infección. La longitud canónica de miARNs maduros se indica con una línea roja discontinua.

La predicción de Tc-mirs con mirDeep2 usando la base de datos completa de los miARNs conocidos hasta la fecha, confirmó la presencia de potenciales miARNs maduros y potencialmente funcionales en células infectadas. Del total de Tc-mirs predichos entre 1 y 6 horas, 59 presentan una conservación completa de la secuencia semilla (y la secuencia parcial de sus precursores) en 86 especies diferentes, incluyendo humano (hsa) y ratón (mmu), cumpliendo así una de las condiciones más importantes para que un ARN corto pueda considerarse un miARN. La conservación de la secuencia semilla es de especial importancia debido a que se considera como la región que reconoce al ARNm diana, de tal manera que los miARNs conservados entre especies son capaces de regular los mismos genes y/o las mismas rutas biológicas[231].

A 24 h.p.i se determinó la existencia 77 potenciales nuevos Tc-mirs, 55 de los cuales presentan una estructura secundaria capaz de auto-estabilizarse formando una doble cadena de ARN (*randfold* significativo). En total, 136 Tc-mirs durante la fase temprana de la infección, cumplen con todos los requisitos de un miARN maduro, estable y potencialmente funcional biológicamente (**Figura 24**).



**Figura 24.** Ejemplo de un miARN de *T. cruzi* predicho con el uso de mirDeep2 (Tc-mir-9-1810) a 24 h.p.i en macrófagos de ratón BALB/c.

A pesar de contar con la evidencia de la presencia de miARNs maduros de *T. cruzi* en el proceso de infección, así como la evidencia *in silico* de su funcionalidad biológica, no fue posible determinar la existencia de la maquinaria molecular indispensable para su biogénesis (en ninguno de los genomas actualmente disponibles del parásito o los dos nuevos presentados en este trabajo) por una ruta canónica o alternativa (*Dicer1* y *Ago2*, principalmente).

Hasta la fecha, se tiene evidencia de que el proceso de producción y uso de miARNs como mecanismo de control de la expresión existe o existió en gran parte de las especies conocidas, sin embargo, en algunas como los parásitos ha desaparecido[68]. Se ha descrito la presencia de pseudo-genes evolutivamente relacionados con genes del tipo argonauta en parásitos causantes de enfermedades humanas como *T. cruzi*, *T. brucei* y *Leishmania* que demuestran la pérdida de este mecanismo en el transcurso de años de evolución[232]. Por lo tanto, con estos

resultados contradictorios proponemos la hipótesis de que la maquinaria del huésped podría estar involucrada en el proceso de maduración de los Tc-mirs del parásito, y que, además, estos Tc-mirs podrían controlar la respuesta inmune del huésped. La activación de *Nod2* y *Tlr7* a 24 h.p.i (detectado por RNAseq) por la presencia de ARN del patógeno en lugar de *Tlr9* y *Nod1* por la eliminación del parásito y el reconocimiento de ADN, es consistente con nuestra hipótesis.

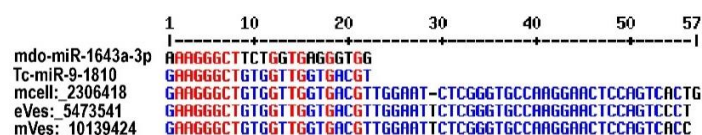
La vía de comunicación huésped-patógeno considerada en nuestra hipótesis la constituyen las vesículas extracelulares de la forma tripomastigote. Desde su descubrimiento en 1979[233], las vesículas extracelulares producidas por *T. cruzi* han ganado atención, se han relacionado en diversos procesos que van desde la internalización del parásito hasta incrementar la carga parasitaria en corazón, además de la evasión de la respuesta inmune[234,235]. Sin embargo, los procesos moleculares que producen estos mecanismos continúan siendo desconocidos. Sorprendentemente, y en consistencia con nuestra hipótesis, se ha propuesto que las vesículas extracelulares de organismos procariotas (de ARN), constituyen un medio de comunicación celular crucial en el desarrollo de enfermedades, mediante la manipulación de la respuesta inmune y la expresión génica[236].

En 2014 Bayer-Santos y col.[237] describieron el contenido de las vesículas extracelulares de *T. cruzi* mediante smallRNAseq (16-50 nucleótidos). Comparó el contenido de las vesículas extracelulares de la forma epimastigote (eVes) con la forma tripomastigote metacíclica (mVes) y el contenido de RNAs cortos de la célula completa (mCell). Sus resultados sugieren que el contenido de las vesículas podría ser dependiente de la cepa de estudio (diferencias entre Dm28c y CI Brener), de la fuente de aislamiento (eVes o mVes) y que contienen hasta 7 tipos de ARN diferente, que van desde el ARNr (el más abundante) hasta ARNpn (el más escaso). El ARN derivado de ARNm (definido como CDS) es el tercero más abundante.

Para determinar si las vesículas extracelulares son el medio de transporte de los Tc-mirs detectados en este trabajo, se usaron los datos de secuenciación masiva Bayer-Santos y col.[237], depositados en el SRA del NCBI con número de acceso SRP035414.

Los resultados muestran que los Tc-mirs definidos en este trabajo se encuentran en las 3 muestras analizadas (eVes, mVes y mCell), en abundancia variable, pero interesantemente, en forma de potenciales miARNs precursores, no en su forma madura (**Figura 25**). Por un lado, indica que las vesículas extracelulares son probablemente el medio por el que el parásito comparte estos ARNs cortos con el huésped, y por otro lado se refuerza la hipótesis de que los Tc-mirs precursores maduran con la maquinaria del huésped.





**Figura 25.** Potenciales precursores del Tc-miR-9-1810 encontrado en las vesículas extracelulares de *T. cruzi* (mCell, eVes, mVes). Este Tc-miR conserva su secuencia semilla con el miR-1643a-3p de *Monodelphis domestica*.

En resumen, se demostró que *T. cruzi* es capaz de intercambiar ARN con el huésped probablemente mediante sus vesículas extracelulares y que, en el huésped se pueden generar miARNs maduros y potencialmente funcionales.

En este punto, el siguiente paso fue demostrar biológicamente que los Tc-mirs actúan en perjuicio de la célula huésped mediante su acción en el RISC. La estrategia diseñada fue la inmunoprecipitación de una de las proteínas más importantes del complejo (*Ago2*). Sin embargo, debido a diversos problemas técnicos, no se obtuvieron resultados significativos que validen que los Tc-mirs sean unidos al complejo de silenciamiento del huésped. Debido a la relevancia de este descubrimiento, se contempla continuar con la validación o rechazo de esta hipótesis en el futuro mediante el uso de técnicas alternativas a la inmunoprecipitación o probando con distintas líneas celulares. Además, se contempla estudiar el efecto de la inhibición de dichos miARNs mediante el uso de antagomirs durante la infección.

#### 4.4 Transcriptómica de corazón de ratón BALB/c en estado crónico de la enfermedad de Chagas

Con el objetivo de ampliar los datos anteriormente presentados y dado que la enfermedad de Chagas se caracteriza clínicamente por una cardiomiopatía en la fase crónica, se realizó un estudio transcriptómico en corazón de animales infectados con distintas cepas de *T. cruzi* que presentan distintos niveles de patogenicidad.

Las cepas de empleadas en este análisis fueron Sylvio X10, Y, CI Brener y VFRA. Un estudio preliminar (RNAseq, 1 muestra por condición con 5 réplicas biológicas) mostró que la cepa VFRA causa una mayor activación de la respuesta inmune humoral, por lo tanto, solo para esta cepa (incluidos controles no infectados) se

secuenciaron las réplicas técnicas (2 adicionales). Aunque los resultados de las diferentes cepas son en general consistentes con estudios anteriores, estos resultados sin réplicas técnicas (Sylvio X10, CI Brener e Y; 5 réplicas biológicas) pueden considerarse como preliminares.

Después del filtrado por calidad, longitud y limpieza de adaptadores de secuenciación, se obtuvieron un total de 63 millones (en promedio) de lecturas por muestra (75 nucleótidos de longitud). En todos los casos, los porcentajes de alineamiento al genoma de ratón fueron superiores al 95%.

Las muestras de cada grupo (control e infectado) mostraron un alto grado de reproducibilidad después de la remoción del efecto *batch* en las muestras de secuenciación y por tanto se continuó con el análisis de expresión diferencial (**Figura Suplementaria 13**).

Los resultados del análisis de expresión diferencial muestran que del transcriptoma total (aproximadamente 22.000 transcritos) hasta 5.700 tienen expresión significativa ( $p$ -valor ajustado  $\leq 0,001$ ), de los cuales, además, 166 se encuentran en niveles significativamente diferentes ( $\text{Fold} \geq 5$ ,  $\leq -5$ ) a la muestra no infectada (**Figura Suplementaria 14**).

De acuerdo con lo esperado, es posible identificar a simple vista algunos de los genes con mayor expresión en la etapa crónica de la enfermedad de Chagas, principalmente asociados a daño cardíaco tales como *ifng*, *Il-2*, *Il27*, *Ccl8*, *Ccl17*, *Cxc19*, *Il6*, entre otros (**Figura Suplementaria 14**).

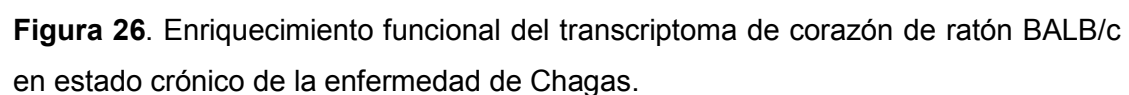
Sin embargo, el análisis de enriquecimiento funcional del transcriptoma completo (genes con expresión diferencial significativa) permitió identificar de manera global y por primera vez en *T. cruzi*, las rutas metabólicas implicadas en el estado crónico de la enfermedad y potencialmente en el daño cardíaco (**Figura 26**).

La fase crónica de la infección por *T. cruzi* se caracteriza principalmente por los elevados niveles de las citoquinas pro-inflamatorias *Il12*, *Il6*, *Il1b*, *Il18*, *Il23* *Ifng*, *Tnfa* entre otras, actuando sinérgicamente para activar a los macrófagos y eliminar al parásito mediante un ambiente de estrés oxidativo dependiente de ácido nítrico y RNI[44,57]. Este repertorio de citoquinas pro-inflamatorias es además responsable de la diferenciación y proliferación de las células  $\text{CD4}^+$  Th y la respuesta  $\text{CD8}^+$ , conteniendo los niveles de parásito, la respuesta inmune y confiriendo inmunidad[63,238]. Por otro lado, algunos sugieren la posibilidad de que el daño cardíaco, en ocasiones letal causado en el estado crónico de la enfermedad o CCC por sus siglas en inglés (*Chronic Chagas Cardiomyopathy*) es originado por la severidad de la respuesta inmune en la etapa temprana y aguda de la infección, que persiste por años con la presencia del parásito[65,239]. Rick L. Tarleton (2007)[57]

sugiere dos escenarios principales involucrados en el desarrollo de este padecimiento, primero, la ineficiente detección del parásito para su eliminación (la respuesta inmune causada por *T. cruzi* es menor a la inducida por LPS-*INF* $\gamma$ )[238,240] y segundo, la inusual capacidad del parásito para evadir el sistema inmune mediante la presentación de una enorme variedad de potenciales epitopos para las células T CD8<sup>+</sup> referida como “confusión inmune” (estudios han demostrado que como máximo, 30% de la totalidad de la población CD8<sup>+</sup> en ratón es específica para un solo péptido)[241].

Además, diversos estudios han demostrado que la respuesta inflamatoria producida principalmente por *IFN* $\gamma$  podría estar asociada con la destrucción del tejido cardíaco, y que con la persistencia puede causar una respuesta autoinmune incontrolable.

Con los resultados actualmente obtenidos para la enfermedad de Chagas (estado crónico) en modelos *in vitro*, *in vivo* y pacientes, se pueden resumir las rutas biológicas implicadas como: Diferenciación celular Th1, Th2, Th17 y Treg y con la producción de las citoquinas *IFN* $\gamma$ , IL2, IL12, TNF $\alpha$  (Th1), Gata3, IL4, IL5, IL6, IL13 (Th2), IL17, IL23 (Th17) y Foxp3, IL10, IL35, Ebi3, TGF $\beta$  (Treg), relacionadas con el proceso inmunidad, daño cardíaco y autoinmunidad (principalmente Th17) y la activación de células NK (*Natural Killer*) aunque el mecanismo de activación y efectores en esta enfermedad son desconocidos[44,57,58,63,242,243].

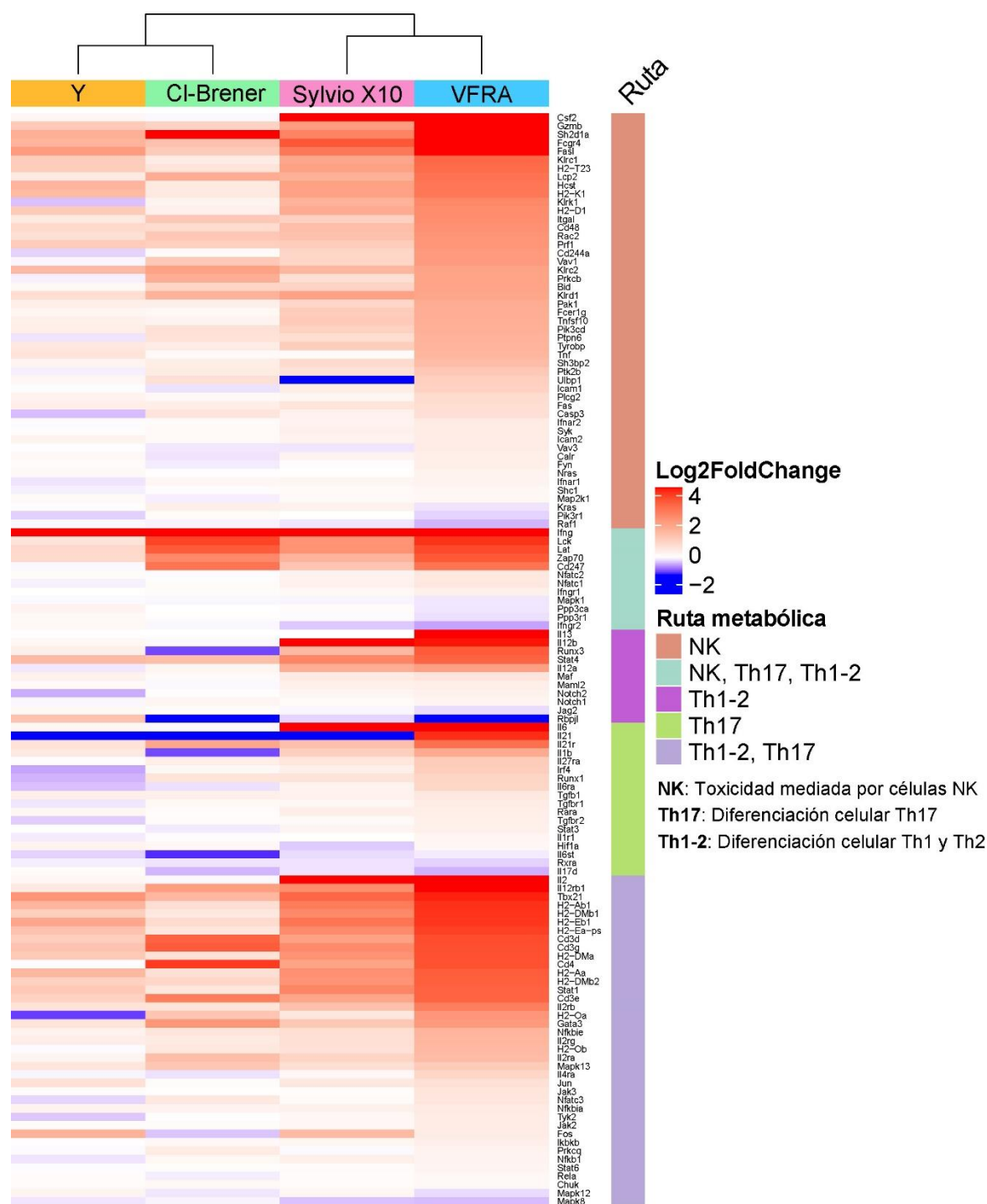


En primer lugar, es posible identificar el enriquecimiento de las rutas esperadas por resultados previos, tales como la misma enfermedad de Chagas (que incluye los genes más analizados durante la enfermedad), la diferenciación de células Th1 y Th2, así como de sus mecanismos de señalización, y complementarios como los mecanismos de procesamiento y presentación de antígenos de las células APC (*Antigen Presenting Cells*) e interferón. Otras relacionadas con la citotoxicidad mediada por células NK, el proceso de endocitosis y respuesta inmune (vía de Nod,

enfermedad inflamatoria intestinal e interacción de citoquinas, principalmente), otras relacionadas con enfermedades por parásitos intracelulares (infección por *Leishmania* y Toxoplasmosis), de forma interesante con miocarditis viral, influenza A, Epstein-Bar, Herpes y finalmente pero no menos interesante, otras rutas relacionadas con enfermedades autoinmunes, como la artritis reumatoide, la diferenciación de las células Th17, tiroiditis autoinmune, diabetes mellitus tipo 1, rechazo a injertos (*Allograft rejection*) y enfermedad de injerto contra el huésped (*Graft-versus-host disease*). Estas dos últimas (*Allograft rejection*, *Graft-versus-host disease*) podrían estar relacionadas con la teoría de la autoinmunidad, mediada por el daño persistente al tejido cardíaco causado por la respuesta inflamatoria enfocada a la eliminación del parásito, que además se conservan en las 3 cepas con mayor afectación del sistema inmune (VFRA, Sylvio X10 e Y; **Tabla Suplementaria 4**).

Aunque en general es posible identificar que las rutas enriquecidas coinciden entre cepas, principalmente para la cepa VFRA y Sylvio X10 (**Tabla Suplementaria 4**), las diferencias principales se encuentran en la cantidad y niveles de expresión de los genes involucrados. En otras palabras, las principales diferencias entre cepas, el nivel de inducción de estas rutas. Tal como sugiere el análisis de componentes principales de las lecturas de secuenciación normalizadas (**Figura Suplementaria 13**) siendo VFRA la cepa que mayor activación de las rutas genera, seguida de Sylvio X10, la cepa Y y por último CI Brener, con un fenotipo muy similar a las muestras no infectadas con solo una ruta activada significativamente (*T-cell receptor signalling pathway*). Estos resultados coinciden con la mayor severidad clínica y presencia del parásito (mayor par VFRA y menor en CI Brener) en la fase crónica de la enfermedad. En ningún caso fue posible detectar rutas significativamente inhibidas (p-valor ajustado  $\leq 0,001$ ).

El análisis individual de las rutas relacionadas con la respuesta inmune e inmunidad (diferenciación celular Th1, Th2, Th17, y toxicidad mediada por células NK) permitió identificar los principales genes con expresión diferencial que determinan el nivel de activación de estos mecanismos biológicos (**Figura 27**).



**Figura 27.** Niveles de expresión (log2Fold respecto a la muestra no infectada) de los genes involucrados en las principales rutas de respuesta inmune e inmunidad enriquecidas funcionalmente durante la infección por *T. cruzi* (estado crónico). Los genes sin variación de expresión respecto a la muestra no infectada (log2Fold = 0) en todas las muestras, no se incluyen en la figura con el objetivo de conseguir una mejor ilustración de las rutas.

Los resultados confirman desequilibrio en células Th en el estado crónico de la enfermedad, siendo Th1 la fracción aparentemente con mayor activación por la sobreexpresión de *Ifng*, *Tbx21* (*Tbet*), *Tnfa*, *Il2* e *Il12* (a y b).

Con la sobreexpresión significativa únicamente de la *Il6* y el factor de transcripción *Gata3*, es posible inferir una baja diferenciación de las células Th2. Lo mismo ocurre para las células Treg con la inhibición de *Foxp3*, *Il10* e *Il35*. Sin embargo, con la sobreexpresión de los genes *Il1b*, *Il21*, *Il17c*, *Stat3*, *Irf4* y *Rara* así como de los receptores *Il17rc*, *Il17rb* e *Il17ra* (principalmente), es posible determinar una elevada activación de la controversial respuesta Th17[46], así como ampliar significativamente el conocimiento de los genes involucrados en su mecanismo de activación en la enfermedad de Chagas. Aunque estos resultados no son suficientes para responder la principal incógnita respecto al rol benéfico o perjudicial de las células Th17 en la fase crónica de la enfermedad, es posible concluir que la cepa VFRA es el mejor modelo para su análisis *in vivo* debido a los niveles de activación que causa en la respuesta humoral y las principales rutas identificadas hasta el momento. Por un lado, se describe por primera vez la sobreexpresión de la *Il17c* en lugar de la *Il17a* asociada ampliamente a miocarditis por la persistencia del parásito[243] y por otro, de forma aún más interesante se vislumbran los dos potenciales mecanismos de su inducción basados en la ruta general diseñada por KEGG (mmu04659); el primero dependiente de la *Il1β*, mapquinasas (*Mapk13*, *Map2k1*) e *Irf4* y otro dependiente del receptor de ácido retinoico (*Rara*).

Con los resultados del presente trabajo y experimentos futuros que determinen la correlación entre los niveles de parásito, la respuesta humoral y los niveles de daño cardíaco podrían resolver la controversial respuesta Th17 en el desarrollo de la enfermedad y su letalidad.

Por otro lado, fue posible la identificación de los principales genes involucrados en la citotoxicidad mediada por células NK (**Figura 28**) desconocidos hasta ahora en la enfermedad de Chagas. Los resultados concuerdan con las suposiciones de Filtjens y col.[244] sobre el papel de los receptores de la familia *Ly49* en la activación de este tipo celular. Sus estudios demostraron que la inhibición del gen *Ly49e* (*Klra5*) se asocia con una mayor actividad antiparasitaria (principalmente IFN $\gamma$ ) como ocurre en la cepa VFRA, sin embargo, fue posible detectar la activación significativa de al menos 3 receptores de la misma familia (*Klra1,3* y 7), sugiriendo que algunos miembros de esta familia genética podrían ser capaces de desarrollar un papel completamente opuesto. Además, fue posible identificar un amplio complejo de receptores de células NK, tales como *Klrk1*, *Klrc1*, *Klrc2* y *Klrd1* con sobreexpresión significativa (**Figura 28**).





Finalmente, las rutas biológicas restantes enriquecidas funcionalmente podrían establecer nuevas estrategias de investigación en la enfermedad de Chagas, principalmente las relacionadas a enfermedades virales (miocarditis viral, virus linfotrópico de células T (virus 1), infección por herpes, VIH y la infección por el virus Epstein-Barr), la relacionadas con daños en otros órganos (enfermedad inflamatoria intestinal) pero de forma más interesante las que podrían justificar retomar seriamente la teoría de daño cardíaco por autoinmunidad (enfermedad tiroidea autoinmune, diabetes mellitus tipo 1, artritis reumatoide, rechazo de aloinjerto y la enfermedad injerto-contrahuésped).

Con hasta 24 genes del complejo de histocompatibilidad I y II o MHC I y MHC II por sus siglas en inglés (*Major Histocompatibility Complex*), diferencial y significativamente expresados y relacionados con el desarrollo de enfermedades autoinmunes[246] (**Tabla suplementaria 5**), se hace necesario diseñar experimentos futuros que complementen el transcriptoma de este trabajo, que permitan confirmar o refutar algunas las hipótesis que esta expresión diferencial genera. Principalmente comprobar si el enriquecimiento funcional de estas rutas se debe a la redundancia de funcionalidad por las diversas funciones en las que estos genes han sido involucrados (principalmente citotoxicidad y diferenciación celular T CD4+ y CD8+) o el rol específico en el desarrollo de una respuesta autodestructiva.

De forma igualmente interesante, es necesario determinar el motivo del enriquecimiento funcional de rutas asociadas a infecciones virales, por una parte, descartando los genes asociados a la respuesta inmune e inmunidad y por otro determinado el motivo de la sobreexpresión significativa de genes relacionados con la detección de material genético exógeno por de modo intracelular, tales como *Tlr7*, *Tlr9*, *Nod1* y *Nod2* en un momento de la infección donde el parásito permanece en niveles mínimos. Además, la sobreexpresión de estos sensores moleculares (principalmente *Nod2* y *Tlr7*) y el enriquecimiento de la ruta de endocitosis (principalmente en Sylvio X10) apoyan la hipótesis de este trabajo sobre la capacidad del parásito de intercambiar material genético con la célula huésped con el objetivo de desarrollar un mecanismo alternativo de evasión de la respuesta inmune y los procesos de inmunidad. Este mecanismo de evasión ventajosa para el parásito, podría estar regulado por la expresión de miARNs, tanto del huésped como del propio patógeno en acuerdo con lo descrito por Rodrigues Pinto Ferreria y col. (2017)[247] cuyos resultados sugieren la participación de miARNs en la alteración de diversas rutas relacionadas con los procesos de inmunidad, respuesta inmune y desarrollo de enfermedades autoinmune descritas en este trabajo, tales como la

diferenciación de células T, respuesta de células NK, la enfermedad injerto-contrahuésped, diabetes mellitus tipo 1 entre otras estrechamente relacionadas.

#### 4.5 Panorama general

El conocimiento de la secuencia genómica completa del parásito es imperativo para su mejor entendimiento a varios niveles, que van desde los mecanismos moleculares que le confieren capacidad de evolución genética constante, hasta las diferencias que desencadenan niveles de virulencia y patogenicidad distintivo entre cepas.

El uso de tecnologías de secuenciación de tercera generación es indispensable para obtener un ensamblaje más completo del genoma de este parásito, principalmente para evitar su fragmentación causado por las regiones de baja complejidad lingüística. El uso de ensamblajes híbridos entre tecnologías de segunda (Illumina) y tercera generación (PacBio/MinION) podría constituir la estrategia ideal.

Al tratarse de un organismo tan complejo, quizá la mejor estrategia sea centrarse en analizar aquellas características que los definen como una sola especie, más que en aquellas que los hacen parecer tan distantes. Primero, el análisis de la composición nucleotídica en las UTRs para determinar si existen motivos específicos que promuevan el reclutamiento de la maquinaria de traducción y que definen la estabilidad de los ARNm. Segundo, el análisis de la composición nucleotídica de los sitios divergentes donde inicia la transcripción, para determinar si existe algún o algunos motivos específicos que promuevan el reclutamiento de la maquinaria molecular. Tercero, con todos los genomas actualmente disponibles en las bases de datos, (ninguno completamente secuenciado) crear un solo genoma consenso de referencia, o pangenoma, como se hace en otras especies (principalmente bacterias y arqueas), estableciendo en su totalidad el genoma conservado (core) y los genes con mayor predisposición a evolución (disruptive). Cuarto, crear un modelo de cultivo *in vitro* e *in vivo* que permita determinar la velocidad con la que se introducen cambios de secuencia significativos cada familia genética.

Por otra parte, los resultados de la expresión de miARNs durante la fase temprana de la infección indican, que 6 h.p.i es el momento de mayor actividad de estos ARNs cortos y que pueden regular algunas de las rutas biológicas más importantes relacionadas con la internalización del parásito, principalmente. Y, por otra parte, uno de los descubrimientos más importantes de este trabajo es la formación (*in silico*) de miARNs del parásito, madurados por la célula huésped y usados potencialmente en su perjuicio. La expresión de los sensores de ARN exógeno (*Tlr7*

y *Nod2*) en la fase temprana de la infección y en la etapa crónica de la enfermedad, la presencia de la secuencia precursora de los Tc-mirs en las vesículas extracelulares, y la conservación de la secuencia semilla (*seed*) de los Tc-mirs teóricos en otras especies, principalmente, respaldan esta teoría. Además, en acuerdo con nuestra teoría, estudios previos de Moreira y col. (2019)[234], determinaron que la incubación de células Vero con vesículas extracelulares de la cepa de *T. cruzi* Pan4 es capaz de producir vulnerabilidad a la internalización del parásito.

Finalmente, el estudio transcriptómico del tejido más afectado en la etapa crónica de la enfermedad, nos ha permitido identificar por primera vez en *T. cruzi* la cepa de referencia más adecuada para estudios *in vivo*, por su mayor virulencia y activación de la respuesta inmune y adquisición de inmunidad por la activación de las células Th1, Th2 y NK.

## 5 CONCLUSIONES

---

## 5 Conclusiones

1.- La secuenciación del genoma de la cepa Bug2148 con la tecnología PacBio permitió obtener la secuencia más larga ensamblada *de novo* en *T. cruzi* hasta la fecha (1.3 Mb), y con esto, identificar las familias proteicas más cercanas a las regiones teloméricas susceptibles a evolución constante.

2.- La reanotación del genoma de la cepa Sylvio X10 mediante datos de su transcriptoma, permitió la anotación de más del 80% de la secuencia total, incrementando considerablemente la información genética de la versión anterior.

3.- La identificación y descripción de los sitios de inserción de SL y poli-A por primera vez en *T. cruzi*, así como la identificación de las principales diferencias a nivel de secuencia con otros parásitos evolutivamente relacionados, constituye un avance importante en el entendimiento de los mecanismos moleculares de maduración del ARNm.

4.- Los resultados del transcriptoma de la cepa Sylvio X10 en estado tripomastigote confirman que algunas de las familias mas abundantes en el genoma, son a su vez las más abundantes a nivel de ARNm, confirmando que el número de copias de cada gen en el genoma consituye un mecanismo adicional en el control de la expresión.

5.- La mayor desregulación en la expresión de miARNs en la fase temprana de la infección ocurre después de 6 horas, e incluye hasta 73 sobre-expresados y 48 inhibidos significativamente.

6.- La inactivación de 29 miARNs a 6 h.p.i y la consecuente sobre-expresión de sus genes diana (hasta 111), puede causar la activación de rutas involucradas en procesos de señalización, reorganización del citoesqueleto, respuesta inmune e interacción huésped patógeno, que convergen en una mayor capacidad de internalización y mayor proliferación intracelular por el parásito.

7.- Análisis *in silico* demuestran que el intercambio de material genético entre el parásito y la célula huésped mediante vesículas extracelulares, podría generar

miARNs del parásito potencialmente funcionales, favoreciendo su capacidad infectiva y la modulación de la respuesta inmune.

8.- El estudio de transcriptómica en corazón de ratón infectado y en estado crónico de la enfermedad de Chagas confirma el tropismo patogénico de las diferentes cepas de *T. cruzi*, siendo VFRA la que genera una mayor activación del sistema inmune.

9.- La activación de sensores intracelulares de material genético exógeno (*Tlr7*, *Tlr9* y *Nod2*) constituye una vía de señalización importante y novedosa en el proceso de activación de respuesta inmune y adquisición de inmunidad.

10.- Por primera vez en *T. cruzi* fue posible determinar que principalmente la activación de las células Th1, Th2 y NK se asocia a diferentes procesos biológicos inherentes a la infección, daño cardíaco y desarrollo del estado crónico de la enfermedad de Chagas.

## 6 BIBLIOGRAFÍA

---

## 6 Bibliografía

1. Chagas C. Nova tripanozomíaze humana. Estudos sobre a morfologia e o ciclo evolutivo de *Schizotrypanum cruzi* n. gen., n. sp., agente etiológico de nova entidade morbida do homem. Mem Inst Oswaldo Cruz. 1909;1: 159–218.
2. WHO | Chagas disease (American trypanosomiasis). In: WHO [Internet]. World Health Organization; 2018 [cited 16 May 2019].
3. Pérez-Molina JA, Molina I. Chagas disease. Lancet. 2018;391: 82–94.
4. Tarleton RL. Chagas Disease: A Solvable Problem, Ignored. Trends Mol Med. 2016;22: 835–838.
5. Schmunis GA, Yadon ZE. Chagas disease: A Latin American health problem becoming a world health problem. Acta Trop. 2010;115: 14–21.
6. Rassi A, Rassi A, Marcondes de Rezende J. American Trypanosomiasis (Chagas Disease). Infect Dis Clin North Am. 2012;26: 275–291. DOI:10.1016/j.idc.2012.03.002: PMID:22632639
7. Rassi A, Marin-Neto A. Seminar Chagas disease [Internet]. Lancet. 2010.
8. Carod-Artal FJ. American trypanosomiasis. Handbook of Clinical Neurology. 1st ed. © 2013, Elsevier B.V. All rights reserved.; 2013. pp. 103–123.
9. Araújo PR, Santuza M, Teixeira /. Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in *Trypanosoma cruzi* -A Review. Mem Inst Oswaldo Cruz. 2011;106: 257–266.
10. Steverding D. The history of Chagas disease [Internet]. 2014.
11. Gómez LA, Gutierrez FRS, Peñuela OA. *Trypanosoma cruzi* infection in transfusion medicine. Hematol Transfus Cell Ther. 2019;
12. Rodrigues Coura J, Borges-Pereira J. Chagas disease: 100 years after its discovery. A systemic review. Acta Trop. 2010;115: 5–13.
13. Apt. B. W, Heitmann G. I, M. Isabel JL, Jofré M. L, Muñoz C. del V. P, H. IN, et al. Tratamiento antiparasitario de la enfermedad de Chagas. Rev Chil infectología. 2008;25: 384–389.
14. Bermudez J, Davies C, Simonazzi A, Pablo Real J, Palma S. Current



- drug therapy and pharmaceutical challenges for Chagas disease. *Acta Trop.* 2016;156: 1–16.
15. Yamagata Y, Nakagawa J. Control of Chagas Disease. 2006. pp. 129–165.
  16. Liu Q, Zhou X-N. Preventing the transmission of American trypanosomiasis and its spread into non-endemic countries. *Infect Dis Poverty.* 2015;4.
  17. Zaniello BA, Kessler DA, Vine KM, Grima KM, Weisenberg SA. Seroprevalence of Chagas Infection in the Donor Population. *PLoS Negl Trop Dis.* 2012;6: 1771.
  18. Gibson W, Bingle L, Blendeman W, Brown J, Wood J, Stevens J. Structure and sequence variation of the trypanosome spliced leader transcript [Internet]. *Molecular and Biochemical Parasitology.* 2000.
  19. Akopyants NS, Kimblin N, Secundino N, Patrick R, Peters N, Lawyer P, et al. Demonstration of genetic exchange during cyclical development of *Leishmania* in the sand fly vector. *Science* (80- ). 2009;10.
  20. Berryid ASF, Salazar-Sá Nchezid R, Castillo-Neyraid R, Borrini-Mayorí K, Chipana-Ramos C, Vargas-Maquera M, et al. Sexual reproduction in a natural *Trypanosoma cruzi* population. *PLoS Negl Trop Dis.* 2019;20.
  21. Souza RT, Bio F, Lima M, Barros RM, Cortez DR, Santos MF, et al. Genome Size, Karyotype Polymorphism and Chromosomal Evolution in *Trypanosoma cruzi*. 2011;
  22. Luís Reis-Cunha J, Rodrigues-Luiz GF, Valdivia HO, Baptista RP, Mendes TAO, Loss De Moraes G, et al. Chromosomal copy number variation reveals differential levels of genomic plasticity in distinct *Trypanosoma cruzi* strains. *BMC Genomics.* 2011;16.
  23. Lima FM, Souza RT, Santori FR, Santos MF, Cortez DR. Interclonal Variations in the Molecular Karyotype of *Trypanosoma cruzi*: Chromosome Rearrangements in a Single Cell-Derived Clone of the G Strain. *PLoS One.* 2013;8: 63738.
  24. Henriksson J, Dujardin JC, Barnabé C, Brisse S, Timperman G, Venegas J, et al. Chromosomal size variation in *Trypanosoma cruzi* is mainly progressive and is evolutionarily informative. *Parasitology.* 2002;124: 277–86.

25. Berná L, Pita S, Chiribao ML, Parodi-Talice A, Alvarez-Valin F, Robello C. Biology of the *Trypanosoma cruzi* Genome. Intech. 2019. p. 13.
26. Lewis MD, Llewellyn MS, Gaunt MW, Yeo M, Carrasco HJ, Miles MA. Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in *Trypanosoma cruzi* populations and expose contrasts between natural and experimental hybrids. Int J Parasitol. 39: 1305–1317.
27. Zingales B, Andrade S, Briones M, Campbell D, Chiari E, Fernandes O, et al. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. Mem Inst Oswaldo Cruz Rio Janeiro. 2009;104: 1051–1054.
28. Marcili A, Lima L, Cavazzana MJ, Junqueira AC V, Veludo HH, Maia F, et al. A new genotype of *Trypanosoma cruzi* associated with bats evidenced by phylogenetic analyses using SSU rDNA, cytochrome b and Histone H2B genes and genotyping based on ITS1 rDNA. Parasitology. 2009;136: 641–655.
29. Barnabé C, Mobarec HI, Jurado MR, Cortez JA, Brenière SF. Reconsideration of the seven discrete typing units within the species *Trypanosoma cruzi*, a new proposal of three reliable mitochondrial clades. Infect Genet Evol. 2016;39: 176–186.
30. Najib M, El-Sayed, Peter J. Myler, Daniella C, Bartholomeu, Daniel Nilsson GA, Anh-Nhi Tran, Elodie Ghedin Elizabeth A, Worthey, Arthur L, Delcher, Gaelle Blandin, Scott J W, Elisabet Caler, Gustavo C, Cerqueira Carole Branche, Brian Haas, Atashi Anupama, Erik Arner LA slund, Philip Attipoe, Esteban Bontempi0 Fre´de´ric Bringaud, Peter Burton, Eithon Cadag, David A C, Mark Carrington, Jonathan Crabtree, Hamid Darban, Jose Franco da Silveira P de J, Kimberly Edwards, Paul T, Englund, Gholam Fazelina, Tamara Feldblyum MF, et al. The Genome Sequence of *Trypanosoma cruzi*, Etiologic Agent of Chagas Disease. Science (80- ). 2005;309: 409–415.
31. Reis-Cunha JL, Bartholomeu DC. *Trypanosoma cruzi* Genome Assemblies: Challenges and Milestones of Assembling a Highly Repetitive and Complex Genome. *T. cruzi* Infection and Protocols. Humana Press, New York, NY; 2019. pp. 1–22.

32. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, Ghedin E, Peacock C, Bartholomeu DC, Haas BJ, Tran AN, Wortman JR, Alsmark UC, Angiuoli S, Anupama A, Badger J, Bringaud F, Cadag E, Carl HN. Comparative Genomics of Trypanosomatid Parasitic Protozoa. *Science* (80- ). 2004;309: 404–409.
33. Weatherly DB, Boehlke C, Tarleton RL. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genomics*. 2009;10.
34. Henriksson J, Dujardin JC, Barnabé C, Brisse S, Timperman G, Venegas J, et al. Chromosomal size variation in *Trypanosoma cruzi* is mainly progressive and is evolutionarily informative. *Parasitology*. 2002;124: 277–86.
35. De Gaudenzi JG, Noe G, Campo VA, Frasch AC, Cassola A. Gene expression regulation in trypanosomatids. *Essays Biochem*. 2011;51: 31–46. DOI:10.1042/bse0510031: PMID:22023440
36. Palenchar JB, Bellofatto V. Gene transcription in trypanosomes. *Mol Biochem Parasitol*. 2006;146: 135–141.
37. Martínez-Calvillo S, Nguyen D, Stuart K, Myler PJ. Transcription Initiation and Termination on *Leishmania major* Chromosome 3. *Eukaryot Cell*. 2004;3: 506–517.
38. Thomas S, Green A, Sturm NR, Campbell DA, Myler PJ. Histone acetylations mark origins of polycistronic transcription in *Leishmania major*. *BMC Genomics*. 2009;10.
39. Rastrojo A, Carrasco-Ramiro F, Martín D, Crespillo A, Reguera RM, Aguado B, et al. The transcriptome of *Leishmania major* in the axenic promastigote stage: transcript annotation and relative expression levels by RNA-seq. *BMC Genomics*. 2013;14.
40. Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C. The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLoS Pathog*. 2010;6: 1–15. DOI:10.1371/journal.ppat.1001090: PMID:20838601
41. Minning TA, Weatherly DB, Atwood Iii J, Orlando R, Tarleton RL. The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*. *BMC Genomics*. 2009;10.

42. Patino LH, Ramírez JD. RNA-seq in kinetoplastids: A powerful tool for the understanding of the biology and host-pathogen interactions. *Infect Genet Evol.* 2017;49: 273–282.
43. De Souza W, Ulisses De Carvalho TM, Santos Barrias E. Review on *Trypanosoma cruzi*: Host Cell Interaction. *Int J Cell Biol.* Hindawi Publishing Corporation; 2010;2010.
44. Rodrigues MM, Oliveira AC, Bellio M. The Immune Response to *Trypanosoma cruzi*: Role of Toll-Like Receptors and Perspectives for Vaccine Development . *J Parasitol Res.* 2012;2012: 1–12.
45. Robello C, Dutra Barbosa J, Rocha D, Ferreira BL, Ferreira ÉR, De Brito M V, et al. BALB/c and C57BL/6 Mice Cytokine Responses to *Trypanosoma cruzi* Infection are Independent of Parasite Strain Infectivity. *Front Microbiol.* 2018;9.
46. Sanoja C, Carbajosa S, Fresno M, Ria Gironè S N. Analysis of the Dynamics of Infiltrating CD4 + T Cell Subsets in the Heart during Experimental *Trypanosoma cruzi* Infection. 2013;
47. Ulisses De Carvalho TM, De Souza W. Early Events Related with the Behaviour of *Trypanosoma cruzi* Within an Endocytic Vacuole in Mouse Peritoneal Macrophages. *CELL Struct Funct.* 1989;14: 383–392.
48. Giordano R, Chammas R, Veiga SS, Colli W, Alves MJ. *Trypanosoma cruzi* binds to laminin in a carbohydrate-independent way. *Brazilian J Med Biol Res-Rev Bras Pesqui medicas e Biol.* 1994;27: 2315–8.
49. Teixeira MMG, Yoshida N. Stage-specific surface antigens of metacyclic trypomastigotes of *Trypanosoma cruzi* identified by monoclonal antibodies. *Mol Biochem Parasitol.* 1986;18: 271–282.
50. Santos Barrias E, Maria Ulisses de Carvalho T, De Souza W, Luis Malchiodi E. *Trypanosoma cruzi*: entry into mammalian host cells and parasitophorous vacuole formation. *Front Immunol.* 2013;4.
51. De Pablos LM, Osuna A. Multigene families in *Trypanosoma cruzi* and their role in infectivity. *Infect Immun.* 2012;80: 2258–2264. DOI:10.1128/IAI.06225-11: PMID:22431647
52. Freitas LM, Lopes S, Santos D, Rodrigues-Luiz GF, Mendes TAO, Rodrigues TS, et al. Genomic Analyses, Gene Expression and Antigenic Profile of the Trans-Sialidase Superfamily of *Trypanosoma*

- cruzi* Reveal an Undetected Level of Complexity. PLoS One. 2011;6: e25914.
53. Santana Silva J, Regina Todeschini A, Freire De Lima G, Susana Leguizamón M, Pascuale CA, Burgos JM, et al. Inactive trans-Sialidase Expression in iTS-null *Trypanosoma cruzi* Generates Virulent Trypomastigotes. Front Cell Infect Microbiol. 2017;7.
  54. Carlos A. Buscaglia, Vanina A. Campo ACCF, and Javier M. Di Noia. *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. Nat Rev. 2006;4: 2935–2947.
  55. Alberto Rodriguez M, Stijlemans B, Gluenz E, da Silva C V, da Silva AA, Teixeira TL, et al. Galectin-3: A Friend but Not a Foe during *Trypanosoma cruzi* Experimental Infection. Front Cell Infect Microbiol | www.frontiersin.org. 2017;7.
  56. Burleigh BA. Host cell signaling and *Trypanosoma cruzi* invasion: do all roads lead to lysosomes? Sci STKE. American Association for the Advancement of Science; 2005;2005: pe36. DOI:10.1126/stke.2932005pe36: PMID:16030288
  57. Tarleton RL. Immune system recognition of *Trypanosoma cruzi*. Curr Opin Immunol. 2007;19: 430–434.
  58. Paiva CN, Medei E, Bozza MT. ROS and *Trypanosoma cruzi*: Fuel to infection, poison to the heart. Plos Pathog. 2018;14.
  59. Machado-Silva A, Cerqueira PG, Grazielle-Silva V, Gadelha FR, Peloso E de F, Teixeira SMR, et al. How *Trypanosoma cruzi* deals with oxidative stress: Antioxidant defence and DNA repair pathways. Mutat Res Mutat Res. 2016;767: 8–22.
  60. Vega-Royero SP, Sibona GJ. Who benefits from cellular immune response during the Chagas disease? Biosystems. 2018;171: 66–73.
  61. Romagnani S. T-cell subsets (Th1 versus Th2). Ann Allergy, Asthma Immunol. Elsevier; 2000;85: 9–21.
  62. Alberto C, Zeraín LF. Inmunología de la infección por *T. cruzi* y de la enfermedad de Chagas Comunicación [Internet]. Enf Emerg. 2006.
  63. Cunha-Neto E, Chevillard C. Chagas disease cardiomyopathy: immunopathology and genetics. Mediators Inflamm. Hindawi; 2014;2014: 683230. DOI:10.1155/2014/683230: PMID:25210230

64. Flávia Nardy A, Freire-de-Lima CG, Morrot A. Immune Evasion Strategies of *Trypanosoma cruzi*. J Immunol Res. Hindawi; 2015;2015: 1–7.
65. Gironès N, Fresno M. Etiology of Chagas disease myocarditis: autoimmunity, parasite persistence, or both? Trends Parasitol. 2003;19: 19–22.
66. Cunha-Neto E, Coelho V, Guilherme L, Fiorelli A, Stolf N, Kalil J. Autoimmunity in Chagas' Disease. Identification of Cardiac Myosin-B13 *Trypanosoma cruzi* Protein Crossreactive T Cell Clones in Heart Lesions of a Chronic Chagas' Cardiomyopathy Patient [Internet]. J. Clin. Invest. 1996.
67. GIRONES N, CARRASCO-MARIN E, CUERVO H, GUERRERO NA, SANOJA C, JOHN S, et al. Role of *Trypanosoma cruzi* Autoreactive T Cells in the Generation of Cardiac Pathology. Ann N Y Acad Sci. John Wiley & Sons, Ltd (10.1111); 2007;1107: 434–444.
68. Zheng Y, Cai X, Bradley JE. microRNAs in parasites and parasite infection. RNA Biol. 2013;10: 371–379.
69. Catalanotto C, Cogoni C, Zardo G. MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions. Int J Mol Sci. 2016;17: 17.
70. Desvignes T, Batzel P, Berezikov E, Eilbeck K, Eppig JT, Mcandrews MS, et al. miRNA Nomenclature: A View Incorporating Genetic Origins, Biosynthetic Pathways, and Sequence Variants. Trends Genet. 2015;31: 613–626.
71. Han J. The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev. 2004;18: 3016–3027.
72. Denli AM, Tops BBJ, Plasterk RHA, Ketting RF, Hannon GJ. Processing of primary microRNAs by the Microprocessor complex. Nature. 2004;432: 231–235. DOI:10.1038/nature03049: PMID:15531879
73. Hammond SM. Dicing and slicing. FEBS Lett. 2005;579: 5822–5829.
74. Yi R, Qin Y, Macara IG, Cullen BR. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. Cold Spring Harbor Laboratory Press; 2003;17: 3011. DOI:10.1101/GAD.1158803: PMID:14681208

75. Militello KT, Refour P, Comeaux CA, Duraisingh MT. Antisense RNA and RNAi in protozoan parasites: Working hard or hardly working? *Mol Biochem Parasitol.* Elsevier; 2008;157: 117–126.
76. Ferreira LRP, Frade AF, Santos RHB, Teixeira PC, Baron MA, Navarro IC, et al. MicroRNAs miR-1, miR-133a, miR-133b, miR-208a and miR-208b are dysregulated in Chronic Chagas disease Cardiomyopathy. *Int J Cardiol.* 2014;175: 409–417.
77. H C Crick WF. Molecular Structure of Nucleic Acids [Internet]. Astbury, W. T., Syrnep. Soc. Exp. Biol. 1, Nucleic Acid. GO (Carob. Univ. Press; 1953.
78. Kchouk M, Gibrat J-F, Elloumi M. Generations of Sequencing Technologies: From First to Next Generation. *Biol Med.* 2017;9.
79. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors (DNA polymerase/nucleotide sequences/bacteriophage 4X174) [Internet]. 1977.
80. Maxam AM, Gilbert W. A new method for sequencing DNA [Internet]. *Biochemistry.* 1977.
81. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics.* 2016;107: 1–8.
82. Qiang-long Z, Shi L, Peng G, Fei-shi L. High-throughput Sequencing Technology and Its Application. *J Northeast Agric Univ (English Ed.* 2014;21: 84–96.
83. Mardis ER. DNA sequencing technologies: 2006–2016. *Nat Protoc.* 2017;12.
84. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, García-Cobos S, et al. Application of next generation sequencing in clinical microbiology and infection prevention. *J Biotechnol.* 2017;243: 16–24.
85. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* 2008;36: 105.
86. Karlsson E, Lärkeryd A, Sjödin A, Forsman M, Stenberg & P. Scaffolding of a bacterial genome using MinION nanopore sequencing OPEN. *Sci Rep.* 2015;5.

87. Yandell M, Ence D. A beginner's guide to eukaryotic genome annotation. *Nat Publ Gr.* 2012;13.
88. Edwards DJ, Holt KE. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microb Inform Exp.* 2013;3.
89. Sohn J, Nam J-W. The present and future of *de novo* whole-genome assembly. *Brief Bioinform.* Narnia; 2016;19: bbw096.
90. Mäkinen V, Salmela L, Ylinen J. Normalized N50 assembly metric using gap-restricted co-linear chaining. *BMC Bioinformatics.* 2012;13.
91. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol.* 2012;19: 455–477.
92. Gusev VD, Nemytikova LA, Chuzhanova NA. On the complexity measures of genetic sequences. *Bioinformatics.* 1999;15: 994–999.
93. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 2019;47: 199–205.
94. Kuleshov M V, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016;44.
95. Cole MB, Risso D, Wagner A, DeTomaso D, Ngai J, Purdom E, et al. Performance Assessment and Selection of Normalization Procedures for Single-Cell RNA-Seq. *Cell Syst. Cell Press;* 2019;8: 315–328.e8.
96. Zhou Y, Zhu J, Tong T, Wang J, Lin B, Zhang J. A statistical normalization method and differential expression analysis for RNA-seq data between different species. *BMC Bioinformatics.* 2019;20.
97. RPKM, FPKM and TPM, clearly explained | RNA-Seq Blog [Internet]. 2015 [cited 10 Jul 2019].
98. Yi H, Raman AT, Zhang H, Allen GI, Liu Z. Detecting hidden batch factors through data-adaptive adjustment for biological effects. *Bioinformatics.* 2018;34: 1141–1147.
99. Calabrese KS, Lagrange PH, Gonçalves Da Costa SC. Chagas' disease: Enhancement of systemic inflammatory reaction in cyclophosphamide treated mice. *Int J Immunopharmacol.* Pergamon;



- 1996;18: 505–514.
100. Souza RT, Lima FM, Barros RM, Cortez DR, Santos MF, Cordero EM, et al. Genome size, karyotype polymorphism and chromosomal evolution in *Trypanosoma cruzi*. PLoS One. 2011;6: e23042. DOI:10.1371/journal.pone.0023042: PMID:21857989
  101. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. [cited 6 Mar 2017].
  102. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. Oxford University Press; 2011;27: 863–4. DOI:10.1093/bioinformatics/btr026: PMID:21278185
  103. Chen-Shan Chin, David H Alexander, Patrick Marks, Aaron A Klammer, James Drake CH, Alicia Clum, Alex Copeland, John Huddleston, Evan E Eichler SWT& JK. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013;10: 563–569.
  104. Biopieces by maasha [Internet]. [cited 6 Mar 2017].
  105. Juan Caballero. SeqComplex by jcaballero [Internet]. [cited 5 Apr 2017].
  106. Fu L, Niu B, Zhu Z, Wu S, Li W. Sequence analysis CD-HIT: accelerated for clustering the next-generation sequencing data. 2012;28: 3150–3152.
  107. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11.
  108. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 2002;30: 1575–1584. DOI:10.1093/nar/30.7.1575: PMID:11917018
  109. James T Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S Lander GG& JPM. Integrative genomics viewer. Nat Biotechnol. 2011;29.
  110. Langmead B, Salzberg L S. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9.
  111. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33.

112. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One*. 2014;9: 112963.
113. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 2008;36: 3420–3435.
114. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: A Sequence Logo Generator. *Genome Res*. 2004;14.
115. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinforma Appl NOTE*. 2009;25: 2078–2079.
116. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol. BioMed Central*; 2009;10.
117. Anders S, Pyl PT, Huber W. HTSeq Python framework to work with high-throughput sequencing data. *Bioinformatics*. Oxford University Press; 2015;31: 166–9. DOI:10.1093/bioinformatics/btu638: PMID:25260700
118. Love MI, Huber W, Anders S. Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biol*. 2011;15.
119. Vlachos IS, Zagganas K, Paraskevopoulou MD, Georgakilas G, Karagkouni D, Vergoulis T, et al. DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Web Serv issue Publ online*. 2015;43.
120. Durinx C, McEntyre J, Appel R, Apweiler R, Barlow M, Blomberg N, et al. Open Peer Review Identifying ELIXIR Core Data Resources [version 2; peer review: 2 approved]. *F1000 Res*. 2017;5.
121. Friedländer MR, MacKowiak SD, Li N, Chen W, Rajewsky N. MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res*. 2012;40: 37–52.
122. Bayer-Santos E, Lima M, Conceição Ruiz J, Almeida IC, Franco Da Silveira J, Paso E. Characterization of the small RNA content of *Trypanosoma cruzi* extracellular vesicles. *Mol Biochem Parasitol*.

- 2014;193: 71–74.
123. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. Oxford University Press; 2014;30: 2114–20. DOI:10.1093/bioinformatics/btu170: PMID:24695404
  124. Pertea M, Kim D, Pertea G, Leek JT, Salzberg SL, Author NP. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie, and Ballgown HHS Public Access Author manuscript. *Nat Protoc*. 2016;11: 1650–1667.
  125. Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res*. Narnia; 2018;46: D661–D667.
  126. Kanehisa Laboratories - Archive [Internet]. [cited 19 Jul 2019].
  127. Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, Teixeira MMG, et al. The revised *Trypanosoma cruzi* subspecific nomenclature: Rationale, epidemiological relevance and research applications. *Infect Genet Evol*. 2012;12: 240–253.
  128. David Ramírez J, Jiménez P, Jaimes J, Poveda C. A systematic review of the *Trypanosoma cruzi* genetic heterogeneity, host immune response and genetic factors as plausible drivers of chronic chagasic cardiomyopathy. *Parasitology*. 2018;146.
  129. Carlier Y, Torrico F. Congenital infection with *Trypanosoma cruzi*: from mechanisms of transmission to strategies for diagnosis and control. *Rev Soc Bras Med Trop*. 2003;36: 767–771.
  130. Calderón J, Maganto-Garcia E, Punzón C, Carrión J, Terhorst C, Fresno M. The Receptor Slamf1 on the Surface of Myeloid Lineage Cells Controls Susceptibility to Infection by *Trypanosoma cruzi*. *PLoS Pathog*. 2012;8: e1002799.
  131. Cuervo H, Pineda MA, Aoki MP, Gea S, Fresno M, Gironès N. Inducible Nitric Oxide Synthase and Arginase Expression in Heart Tissue during Acute *Trypanosoma cruzi* Infection in Mice: Arginase I Is Expressed in Infiltrating CD68. *J Infect Dis*. 2008;197: 1772–1782.
  132. Cuervo H, Guerrero NA, Carbajosa S, Beschin A, De Baetselier P, Gironès, Núria and Fresno M. Myeloid-Derived Suppressor Cells

- Infiltrate Myeloid-Derived Suppressor Cells Infiltrate the Heart in Acute *Trypanosoma cruzi* Infection. J Immunol. 2011;187: 2656–2665.
133. Arner E, Kindlund E, Nilsson D, Farzana F, Ferella M, Tammi MT, et al. Database of *Trypanosoma cruzi* repeated genes: 20,000 additional gene variants. BMC Genomics. 2007;8. DOI:10.1186/1471-2164-8-391: PMID:17963481
  134. Franzén O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, Miles MA, et al. Shotgun sequencing analysis of *Trypanosoma cruzi* Sylvio X10/1 and comparison with *T. cruzi* VI CL Brener. PLoS Negl Trop Dis. 2011;5: 1–9. DOI:10.1371/journal.pntd.0000984: PMID:21408126
  135. Oliveira IA, Freire-de-Lima L, Penha LL, Dias WB, Todeschini AR. *Trypanosoma cruzi* Trans-Sialidase: Structural Features and Biological Implications. In: Santos ALS, Branquinha MH, d'Avila-Levy CM, Kneipp LF, Sodré CL, editors. Proteins and Proteomics of Leishmania and Trypanosoma. Dordrecht: Springer Netherlands; 2014. pp. 181–201.
  136. Lander N, Bernal C, Diez N, Añez N, Docampo R, Luis Ramírez J. Localization and Developmental Regulation of a Dispersed Gene Family 1 Protein in *Trypanosoma cruzi*. Infect Immun. 2010;78: 231–240.
  137. Kawashita SY, Da Silva C V, Mortara RA, Burleigh BA, Briones MRS. Homology, paralogy and function of DGF-1, a highly dispersed *Trypanosoma cruzi* specific gene family and its implications for information entropy of its encoded proteins. Mol Biochem Parasitol. 2009;165: 19–31.
  138. Schenkman S, Eichinger D, Pereira MEA, Nussenzweig V. Structural and functional properties of trypanosoma trans-sialidase. Annu Rev Microbiol. 1994;48: 499–523.
  139. A.C.C. Frasch. Functional Diversity in the Trans-sialidase and Mucin Families in *Trypanosoma cruzi*. Parasitol Today. 2000;16: 282–286. DOI:https://doi.org/10.1016/S0169-4758(00)01698-7: PMID:10858646
  140. Chiurillo MA, Cortez DR, Lima FM, Cortez C, Ramírez JL, Martins AG, et al. The diversity and expansion of the trans-sialidase gene family is a common feature in *Trypanosoma cruzi* clade members. MEEGID. 2016;37: 266–274.
  141. Nardy AFFR, Freire-de-Lima CG, Pérez AR, Morrot A. Role of

- Trypanosoma cruzi* Trans-sialidase on the escape from host immune surveillance. *Front Microbiol.* 2016;7. DOI:10.3389/fmicb.2016.00348: PMID:27047464
142. Cremona ML, Campetella O, Sánchez DO, Frasch AC. Enzymically inactive members of the trans-sialidase family from *Trypanosoma cruzi* display beta-galactose binding activity. *Glycobiology.* 1999;9: 581–587.
  143. Mathieu-Daudé F, Lafay B, Touzet O, Lelièvre J, Parrado F, Bosseno M-F, et al. Exploring the FL-160-CRP gene family through sequence variability of the complement regulatory protein (CRP) expressed by the trypomastigote stage of *Trypanosoma cruzi*. *Infect Genet Evol.* 2008;8: 258–266.
  144. Ramírez-Tolosa G, Ferreira A. *Trypanosoma cruzi* evades the complement system as an efficient strategy to survive in the mammalian host: The specific roles of host/parasite molecules and *Trypanosoma cruzi* calreticulin. *Front Microbiol.* 2017;8. DOI:10.3389/fmicb.2017.01667: PMID:28919885
  145. Ramírez JD, Tapia-Calle G, Muñoz-Cruz G, Poveda C, Rendón LM, Hincapié E, et al. Trypanosome species in neo-tropical bats: Biological, evolutionary and epidemiological implications. *Infect Genet Evol.* 2014;22: 250–256.
  146. Kelly S, Carrington M, Kramer S, Schwede A, Maini PK, Gull K. Genome organization is a major component of gene expression control in response to stress and during the cell division cycle in trypanosomes.
  147. Clayton C. The Regulation of Trypanosome Gene Expression by RNA-Binding Proteins Trypanosomes Depend on Post-transcriptional Mechanisms to Regulate Gene Expression. *Plos Pathog.* 2013;9.
  148. Reis-Cunha JL, Rodrigues-Luiz GF, Valdivia HO, Baptista RP, Mendes TAO, de Moraes GL, et al. Chromosomal copy number variation reveals differential levels of genomic plasticity in distinct *Trypanosoma cruzi* strains. *BMC Genomics.* 2015;16. DOI:10.1186/s12864-015-1680-4: PMID:26141959
  149. Dirney Alba Soto C, Garat B, Biochemistry C, Paolo DA, Pastro L, Smircich P, et al. Nuclear Compartmentalization Contributes to Stage-Specific Gene Expression Control in *Trypanosoma cruzi*. *Front Cell Dev*

- Biol | www.frontiersin.org. 2017;5: 8.
150. Vargas N, Pedroso A, Zingales B. Chromosomal polymorphism, gene synteny and genome size in *T. cruzi* I and *T. cruzi* II groups. Mol Biochem Parasitol. 2004;138: 131–141.
  151. Kim BD, Chiurillo MA, El-Sayed N, Jones K, Santos MRM, Porcile PE, et al. Telomere and subtelomere of *Trypanosoma cruzi* chromosomes are enriched in (pseudo)genes of retrotransposon hot spot and trans-sialidase-like gene families: the origins of *T. cruzi* telomeres. Gene. 2005;346: 153–161.
  152. Moraes Barros RR, Marini MM, Antônio CR, Cortez DR, Miyake AM, Lima FM, et al. Anatomy and evolution of telomeric and subtelomeric regions in the human protozoan parasite *Trypanosoma cruzi*. BMC Genomics. 2012;13.
  153. Vazquez MP. The Genetics and Genomics of *Trypanosoma cruzi*. Int J Biomed Pharm Sci. 2007;1: 1–11.
  154. Kudla G, Helwak A, Lipinski L. Gene Conversion and GC-Content Evolution in Mammalian Hsp70. Mol Biol Evol. 2004;21.
  155. Galtier N. Gene conversion drives GC content evolution in mammalian histones. Trends Genet. 2003;19: 65–68.
  156. Dong Kim, Miguel Angel Chiurillo, Najib El-Sayed, Kristin Jones, Marcia R.M. Santos, Patricio E. Porcile, Bjorn Andersson, Peter Mylerf, Jose Franco da Silveira JLR. Telomere and subtelomere of *Trypanosoma cruzi* chromosomes are enriched in (pseudo)genes of retrotransposon hot spot and trans-sialidase-like gene families: the origins of *T. cruzi* telomeres. Gene. 2005;14: 153–161.
  157. Miguel Angel Chiurillo, Isabel Cano JFDS, Jose Luis Ramirez. Organization of telomeric and sub-telomeric regions of chromosomes from the protozoan parasite *Trypanosoma cruzi*. Mol Biochem Parasitol. 1999;100: 173–183.
  158. Turowec JP, Duncan JS, French AC, Gyenis L, Denis NAS, Vilck G, et al. Chapter 23 - Protein Kinase CK2 Is a Constitutively Active Enzyme that Promotes Cell Survival: Strategies to Identify CK2 Substrates and Manipulate its Activity in Mammalian Cells. Methods Enzymol Const Act Recept Other Proteins, Part A. 2010;484: 471–493.

159. Bartholomeu DC, Cerqueira GC, Carolina A, Leã O A, Darocha WD, Pais FS, et al. Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen *Trypanosoma cruzi*. *Nucleic Acids Res.* 2009;37: 3407–3417.
160. Picchi GFA, Zulkiewicz V, Krieger MA, Zanchin NT, Goldenberg S, De Godoy LMF. Post-translational Modifications of *Trypanosoma cruzi* Canonical and Variant Histones. *J Proteome Res.* 2017;16: 1167–1179.
161. Kordyukova MY, Kalmykova AI. Nature and Functions of Telomeric Transcripts. *Biochemistry.* 2019;84.
162. Tibbetts RS, Jensen JL, Olson CL, Wang FD, Engman DM. The DnaJ family of protein chaperones in *Trypanosoma cruzi*: Nucleotide sequence data reported in this paper have been submitted to the GenBank™ data base with the following accession numbers: tcj1, L42541; tcj2, L42549; tcj3, L46818; tcj4, L46819.1. *Mol Biochem Parasitol.* 1998;91: 319–326.
163. Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M. High Guanine and Cytosine Content Increases mRNA Levels in Mammalian Cells. *Plos Biol.* 2006;4: e180.
164. Díaz-Viraqué F, Pita S, Greif G, De Cássia Moreira De Souza R, Iraola G, Robello C. Nanopore sequencing significantly improves genome assembly of the protozoan parasite. *Mol Biol Evol.* 2019;
165. Eranthie Weerapana and Barbara Imperiali. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology.* 2006;16: 91–101.
166. molybdopterin cofactor biosynthetic process Gene Ontology Term (GO:0032324) [Internet]. [cited 23 Nov 2017].
167. Avila JL, Avila A. *Trypanosoma cruzi*: Allopurinol in the Treatment of Mice with Experimental Acute Chagas Disease. *Exp Parasitol.* 1981;51: 204–208.
168. Queiroz R, Benz C, Fellenberg K, Hoheisel JD, Clayton C. Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. 2009;
169. Robello C, Berná L, Chiribao ML, Greif G, Rodriguez M, Alvarez-Valin F. Transcriptomic analysis reveals metabolic switches and surface

- remodeling as key processes for stage transition in *Trypanosoma cruzi*. PeerJ. 2017;5.
170. Callejas-Hernández F, Rastrojo A, Poveda C, Gironès N, Fresno M. Genomic assemblies of newly sequenced *Trypanosoma cruzi* strains reveal new genomic expansion and greater complexity. Sci Rep. Nature Publishing Group; 2018;8: 14631.
  171. Jäger A V, De Gaudenzi JG, Cassola A, Frasch AC. mRNA maturation by two-step trans-splicing/ polyadenylation processing in trypanosomes. Natl Acad Sci. 2000;104: 2035–2042.
  172. Borst P. Discontinuous transcription and antigenic variation in trypanosomes. Ann Rev Biochem. 1986;55: 701–732.
  173. Mccarthy-Burke C, Taylor ZA, Buck GA. Characterization of the spliced leader genes and transcripts in *Trypanosoma cruzi*\* (Chagas' disease; trans-splicing; mRNA processing; RNA secondary structure SL gene, SL-RNA; Y-branch structure) [Internet]. Gene. 1989.
  174. Boothroyd J., Cross GA. Transcripts coding for variant surface glycoproteins of *Trypanosoma brucei* have a short, identical exon at their 5' end. Gene. 1982;20: 281–289.
  175. Pedersen L, Andersen GR, Knudsen CR, Goss Kinzy T, Nyborg J. Biological Crystallography Crystallization of the yeast elongation factor complex. Acta Cryst. 2001;57: 159–161.
  176. Ronalte Alves L, Oliveira C, Goldenberg S. Eukaryotic translation elongation factor-1 alpha is associated with a specific subset of mRNAs in *Trypanosoma cruzi*. BMC Microbiol. 2015;15.
  177. Brandau S, Dresel A, Clos J. High constitutive levels of heat-shock proteins in human-pathogenic parasites of the genus *Leishmania*. Biochem J. 1995;310: 225–232.
  178. Umaer K, Williams N. Kinetoplastid Specific RNA-Protein Interactions in *Trypanosoma cruzi* Ribosome Biogenesis. PLoS One. 2015;10.
  179. Helena Wippel H, Haruo Inoue A, Medeiros Vidal N, Ferreira da Costa J, Hilzendeger Marcon B, Accioly Alves Romagnoli B, et al. Assessing the partners of the RBP9-mRNP complex in *Trypanosoma cruzi* using shotgun proteomics and RNA-seq Assessing the partners of the RBP9-mRNP complex in *Trypanosoma cruzi* using shotgun proteomics and



- RNA-seq. RNA Biol. 2018;15.
180. Patel GP, Ma S, Bag J. The autoregulatory translational control element of poly(A)-binding protein mRNA forms a heteromeric ribonucleoprotein complex. Nucleic Acids Res. 2005;33: 7074–7089.
  181. Kamina AD, Williams N. Ribosome Assembly in Trypanosomatids: A Novel Therapeutic Target. Trends Parasitol. 2017;33: 256–257.
  182. Liu Z, Gutierrez-Vargas C, Wei J, Grassucci RA, Ramesh M, Espina N, et al. Structure and assembly model for the *Trypanosoma cruzi* 60S ribosomal subunit. PNAS. 2016;113: 12174–12179.
  183. Brito Querido J, Mancera-Martínez E, Vicens Q, Bochler A, Chicher J, Simonetti A, et al. The cryo-EM Structure of a Novel 40S Kinetoplastid-Specific Ribosomal Protein. Structure. 2017;25: 1785–1794.e3.
  184. Martínez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ. Transcription of *Leishmania major* Friedlin Chromosome 1 Initiates in Both Directions within a Single Region. Mol Cell. 2003;11.
  185. Li D, Roberts R. WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. Cell Mol Life Sci. 2001;58: 2085–97.
  186. Schapira M, Tyers M, Torrent M, Arrowsmith CH. WD-repeat domain proteins: a novel target class? Nat Rev Drug Discov. 2017;16: 773–786.
  187. Leipe DD, Wolf YI, Koonin E V, Aravind L. Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol. 2002;317: 41–72. DOI:10.1006/jmbi.2001.5378: PMID:11916378
  188. Concanavalin A-like lectin/glucanase domain superfamily (IPR013320) & InterPro & EMBL-EBI [Internet]. [cited 18 Mar 2019].
  189. Günzl A. The Pre-mRNA Splicing Machinery of Trypanosomes: Complex or Simplified? Eukaryot Cell. 2010;9: 1159–1170.
  190. Mair G, Shi H, Li H, Djikeng A, Aviles HO, Bishop JR, et al. A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. RNA. 2000;6.
  191. Dillon LAL, Okrah K, Hughitt VK, Suresh R, Li Y, Fernandes MC, et al. Transcriptomic profiling of gene expression and RNA processing during *Leishmania major* differentiation. Nucleic Acids Res. 2015;43: 6799–6813.

192. Xu P, Wen L, Benegal G, Wang X, Buck GA. Identification of a spliced leader RNA binding protein from *Trypanosoma cruzi* [Internet]. Molecular & Biochemical Parasitology. 2001.
193. Huang J, Van Der Ploeg' LHT. Requirement of a polypyrimidine tract for trans-splicing in trypanosomes: discriminating the PARP promoter from the immediately adjacent 3' splice acceptor site. EMBO J. 1991;10: 3877–3885.
194. Beaudoin E, Freier S, Wyatt JR, Claverie J-M, Gautheret D. Patterns of Variant Polyadenylation Signal Usage in Human Genes. Genome Res. 2000;10: 1001–1010.
195. Tian B, Hu J, Zhang H, Lutz CS. A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res. 2005;33: 201–212.
196. Wahle E, Keller W. The biochemistry of 3'-end cleavage and polyadenylation of messengers rna precursors. Annu Rev Biochem. 1992;61: 419–440.
197. Poveda Cuevas C. Influence of *Trypanosoma cruzi* and host genetic variability on Chagas disease immunopathology [Internet]. Universidad Autónoma de Madrid. 2016.
198. Su S, Zhao Q, He C, Huang D, Liu J, Chen F, et al. ARTICLE miR-142-5p and miR-130a-3p are regulated by IL-4 and IL-13 and control profibrogenic macrophage program. Nat Commun. 2015;6.
199. Adesse D, Goldenberg RC, Fortes FS, Jasmin, Iacobas DA, Iacobas S, et al. Gap Junctions and Chagas Disease. Adv Parasitol. Academic Press; 2011;76: 63–81.
200. Vilar-Pereira G, Silva AA da, Pereira IR, Silva RR, Moreira OC, de Almeida LR, et al. *Trypanosoma cruzi*-induced depressive-like behavior is independent of meningoencephalitis but responsive to parasiticide and TNF-targeted therapeutic interventions. Brain Behav Immun. Academic Press; 2012;26: 1136–1149.
201. Bellera CL, Balcazar DE, Alberca L, Labriola CA, Talevi A, Carrillo C. Identification of levothyroxine antichagasic activity through computer-aided drug repurposing. ScientificWorldJournal. Hindawi; 2014;2014: 279618. DOI:10.1155/2014/279618: PMID:24592161

202. Nde PN, Lima MF, Johnson CA, Pratap S, Villalta F, Mortara RA. Regulation and use of the extracellular matrix by *Trypanosoma cruzi* during early infection. *Front Immunol.* 2012;3.
203. Nde PN, Simmons KJ, Kleshchenko YY, Pratap S, Lima MF, Villalta F. Silencing of the Laminin  $\gamma$ -1 Gene Blocks *Trypanosoma cruzi* Infection. *Infect Immun. American Society for Microbiology (ASM);* 2006;74: 1643. DOI:10.1128/IAI.74.3.1643-1648.2006: PMID:16495535
204. Simmons KJ, Nde PN, Kleshchenko YY, Lima MF, Villalta F. Stable RNA interference of host thrombospondin-1 blocks *Trypanosoma cruzi* infection. *FEBS Lett.* 2006;580: 2365–2370. DOI:10.1016/j.febslet.2006.03.054: PMID:16616140
205. Leal Cunha-e-Silva N, Sant’Anna C, G. Pereira M, De Souza W. Endocytosis in *Trypanosoma cruzi*. *open Parasitol.* 2010;4.
206. Ferreira BL, Ferreira ÉR, Bonfim-Melo A, Mortara RA, Bahia D. *Trypanosoma cruzi* extracellular amastigotes selectively trigger the PI3K/Akt and Erk pathways during HeLa cell invasion. *Microbes Infect. Elsevier Masson;* 2019;
207. Ferrão PM, d’Avila-Levy CM, Araujo-Jorge TC, Degraive WM, Gonçalves A da S, Garzoni LR, et al. Cruzipain Activates Latent TGF- $\beta$  from Host Cells during *T. cruzi* Invasion. Tanowitz HB, editor. *PLoS One.* 2015;10: e0124832. DOI:10.1371/journal.pone.0124832: PMID:25938232
208. de Melo TG, Meirelles M de NSL, Pereira MCS. *Trypanosoma cruzi* alters adherens junctions in cardiomyocytes. *Microbes Infect.* 2008;10: 1405–1410. DOI:10.1016/j.micinf.2008.07.044: PMID:18778790
209. Onofre TS, Rodrigues JPF, Yoshida N. Depletion of Host Cell Focal Adhesion Kinase Increases the Susceptibility to Invasion by *Trypanosoma cruzi* Metacyclic Forms. *Front Cell Infect Microbiol.* 2019;9: 231. DOI:10.3389/fcimb.2019.00231: PMID:31297342
210. Calvet CM, Silva TA, De Melo TG, De Araújo-Jorge TC, Pereira MCDS. TGF-  $\beta$  receptor type II costameric localization in cardiomyocytes and host cell TGF-  $\beta$  response is disrupted by *Trypanosoma cruzi* infection. *Parasitology. Cambridge University Press;* 2016;143: 704–715.
211. Reignault LC, Alcantara C de L, Barrias ES, de Souza W. 3D

- reconstruction of *Trypanosoma cruzi*-macrophage interaction shows the recruitment of host cell organelles towards parasitophorous vacuoles during its biogenesis. *J Struct Biol. Academic Press*; 2019;205: 133–146.
212. Maganto-Garcia E, Punzon C, Terhorst C, Fresno M. Rab5 Activation by Toll-Like Receptor 2 is Required for *Trypanosoma cruzi* Internalization and Replication in Macrophages. *Traffic. John Wiley & Sons, Ltd* (10.1111); 2008;9: 1299–1315.
  213. Bhui T, Roy JK. Rab proteins: The key regulators of intracellular vesicle transport. *Exp Cell Res. Academic Press*; 2014;328: 1–19.
  214. Hovsepian E, Penas F, Siffo S, Mirkin GA, Goren NB, Rottenberg ME, et al. IL-10 Inhibits the NF- $\kappa$ B and ERK/MAPK-Mediated Production of Pro-Inflammatory Mediators by Up-Regulation of SOCS-3 in *Trypanosoma cruzi*-Infected Cardiomyocytes. *Trypanos cruzi-Infected Cardiomyocytes PLoS ONE*. 2013;8: 79445.
  215. Stahl P, Ruppert V, Schwarz RT, Meyer T. *Trypanosoma cruzi* Evades the Protective Role of Interferon-Gamma-Signaling in Parasite-Infected Cells. Rottenberg ME, editor. *PLoS One*. 2014;9: e110512. DOI:10.1371/journal.pone.0110512; PMID:25340519
  216. Lu L, McCurdy S, Huang S, Zhu X, Peplowska K, Tiirikainen M, et al. Time Series miRNA-mRNA integrated analysis reveals critical miRNAs and targets in macrophage polarization. *Sci Rep. Nature Publishing Group*; 2016;6. DOI:10.1038/srep37446; PMID:27981970
  217. ZHANG Y, ZHANG M, ZHONG M, SUO Q, LV K. Expression profiles of miRNAs in polarized macrophages. *Int J Mol Med. Spandidos Publications*; 2013;31: 797–802.
  218. Shi H, Dong G, Yan F, Zhang H, Li C, Ma Q, et al. Arctigenin Ameliorates Inflammation by Regulating Accumulation and Functional Activity of MDSCs in Endotoxin Shock. *Inflammation. Springer US*; 2018;41: 2090–2100.
  219. Aungier SR, Ohmori H, Clinton M, Mabbott NA. MicroRNA-100-5p indirectly modulates the expression of Il6, Ptgs1/2 and Tlr4 mRNA in the mouse follicular dendritic cell-like cell line, FL-Y. *Immunology. Wiley-Blackwell*; 2015;144: 34–44. DOI:10.1111/imm.12342; PMID:24944008

220. Marques-Rocha JL, Garcia-Lacarte M, Samblas M, Bressan J, Martínez JA, Milagro FI. Regulatory roles of miR-155 and let-7b on the expression of inflammation-related genes in THP-1 cells: effects of fatty acids. *J Physiol Biochem.* 2018;74: 579–589. DOI:10.1007/s13105-018-0629-x; PMID:29790117
221. Ana-Carolina Oliveira, Jaqueline R. Peixoto, Luciana B. de Arruda, Marco A. Campos, Ricardo T. Gazzinelli, Douglas T. Golenbock, et al. Expression of Functional TLR4 Confers Proinflammatory Responsiveness to *Trypanosoma cruzi* Glycoinositolphospholipids and Higher Resistance to Infection with *T. cruzi*. *J Immunol.* 2004;173: 5688–5696.
222. Oliveira A-C, de Alencar BC, Tzelepis F, Klezewsky W, da Silva RN, Neves FS, et al. Impaired Innate Immunity in Tlr4<sup>-/-</sup> Mice but Preserved CD8<sup>+</sup> T Cell Responses against *Trypanosoma cruzi* in Tlr4<sup>-</sup>, Tlr2<sup>-</sup>, Tlr9<sup>-</sup> or Myd88-Deficient Mice. Müller I, editor. *PLoS Pathog.* Public Library of Science; 2010;6: e1000870.
223. Huang F, Zhao J-L, Wang L, Gao C-C, Liang S-Q, An D-J, et al. miR-148a-3p Mediates Notch Signaling to Promote the Differentiation and M1 Activation of Macrophages. *Front Immunol.* Frontiers; 2017;8: 1327.
224. Chi X, Ding B, Zhang L, Zhang J, Wang J, Zhang W. lncRNA GAS5 promotes M1 macrophage polarization via miR-455-5p/SOCS3 pathway in childhood pneumonia. *J Cell Physiol.* John Wiley & Sons, Ltd; 2019;234: 13242–13251.
225. Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, Vlachos IS, Tastsoglou S, Kanellos I, et al. DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions. *Nucleic Acids Res.* Narnia; 2018;46: D239–D245.
226. Silva GK, Gutierrez FRS, Guedes PMM, Horta C V, Cunha LD, Mineo TWP, et al. Cutting edge: nucleotide-binding oligomerization domain 1-dependent responses account for murine resistance against *Trypanosoma cruzi* infection. *J Immunol.* American Association of Immunologists; 2010;184: 1148–52. DOI:10.4049/jimmunol.0902254; PMID:20042586
227. Caetano BC, Carmo BB, Melo MB, Cerny A, dos Santos SL,

- Bartholomeu DC, et al. Requirement of UNC93B1 Reveals a Critical Role for TLR7 in Host Resistance to Primary Infection with *Trypanosoma cruzi*. J Immunol. 2011;187: 1903–1911. DOI:10.4049/jimmunol.1003911: PMID:21753151
228. Papathanassiou AE, Ko J-H, Imprialou M, Bagnati M, Srivastava PK, Vu HA, et al. BCAT1 controls metabolic reprogramming in activated human macrophages and is associated with inflammatory diseases. Nat Commun. 2017;8: 16040. DOI:10.1038/ncomms16040: PMID:28699638
  229. O'Neill LAJ, Pearce EJ. Immunometabolism governs dendritic cell and macrophage function. J Exp Med. Rockefeller University Press; 2016;213: 15–23. DOI:10.1084/JEM.20151570: PMID:26694970
  230. Calderon-Gonzalez R, Frande-Cabanes E, Teran-Navarro H, Marimon JM, Freire J, Salcines-Cuevas D, et al. GNP-GAPDH; nanovaccines prevent neonatal listeriosis by blocking microglial apoptosis and bacterial dissemination. Oncotarget. 2017;8: 53916–53934. DOI:10.18632/oncotarget.19405: PMID:28903312
  231. Lewis BP, Shih I, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of Mammalian MicroRNA Targets. Cell. 2003;115: 787–798.
  232. Garcia Silva MR, Tosar JP, Frugier M, Pantano S, Bonilla B, Esteban L, et al. Cloning, characterization and subcellular localization of a *Trypanosoma cruzi* argonaute protein defining a new subfamily distinctive of trypanosomatids. Gene. Elsevier; 2010;466: 26–35.
  233. Da Silveira JF, Abrahamsohn PA, Colli W. Plasma membrane vesicles isolated from epimastigote forms of *Trypanosoma cruzi*. Biochim Biophys Acta - Biomembr. 1979;550: 222–232. DOI:10.1016/0005-2736(79)90209-8: PMID:365244
  234. Moreira LR, Serrano FR, Osuna Id A. Extracellular vesicles of *Trypanosoma cruzi* tissue-culture cell-derived trypomastigotes: Induction of physiological changes in non-parasitized culture cells. PLoS Negl Trop Dis. 2019;13.
  235. Marcilla A, Martin-Jaular L, Trelis M, de Menezes-Neto A, Osuna A, Bernal D, et al. Extracellular vesicles in parasitic diseases. J Extracell vesicles. Taylor & Francis; 2014;3: 25040. DOI:10.3402/jev.v3.25040: PMID:25536932

236. Tsatsaronis JA, Franch-Arroyo S, Resch U, Charpentier E. Extracellular Vesicle RNA: A Universal Mediator of Microbial Communication? Trends Microbiol. Elsevier Current Trends; 2018;26: 401–410.
237. Bayer-Santos E, Lima FM, Ruiz JC, Almeida IC, da Silveira JF. Characterization of the small RNA content of *Trypanosoma cruzi* extracellular vesicles. Mol Biochem Parasitol. 2014;193: 71–74.
238. Bonney KM, Luthringer DJ, Kim SA, Garg NJ, Engman DM. Pathology and Pathogenesis of Chagas Heart Disease. Annu Rev Pathol Mech Dis. Annual Reviews; 2019;14: 421–447.
239. Fresno M, Gironès N. Regulatory Lymphoid and Myeloid Cells Determine the Cardiac Immunopathogenesis of *Trypanosoma cruzi* Infection. Front Microbiol. Frontiers Media SA; 2018;9. DOI:10.3389/fmicb.2018.00351: PMID:29545782
240. Koo S-J, Chowdhury IH, Szczesny B, Wan X, Garg NJ. Macrophages Promote Oxidative Metabolism To Drive Nitric Oxide Generation in Response to *Trypanosoma cruzi*. Infect Immun. 2016;84.
241. Martin DL, Weatherly DB, Laucella SA, Cabinian MA, Crim MT, Sullivan S, et al. CD8+ T-Cell Responses to *Trypanosoma cruzi* are Highly Focused on Strain-Variant trans-Sialidase Epitopes. PLoS Pathog. 2006;2: e77. DOI:10.1371/journal.ppat.0020077: PMID:16879036
242. Geraldo Freire De Lima C, Hölscher C, Gollob KJ, Santana Silva J, Cláudia Silva M, Silva Medina T, et al. Ebi3 Prevents *Trypanosoma cruzi*-Induced Myocarditis by Dampening IFN- $\gamma$ -Driven Inflammation. Front Immunol. 2017;8.
243. Guedes PM da M, Gutierrez FRS, Maia FL, Milanezi CM, Silva GK, Pavanelli WR, et al. IL-17 Produced during *Trypanosoma cruzi* Infection Plays a Central Role in Regulating Parasite-Induced Myocarditis. Correa-Oliveira R, editor. PLoS Negl Trop Dis. Public Library of Science; 2010;4: e604.
244. Filtjens J, Coltel N, Cencig S, Taveirne S, Van Ammel E, Van Acker A, et al. The Ly49E Receptor Inhibits the Immune Control of Acute *Trypanosoma cruzi* Infection. Front Immunol. 2016;10. DOI:10.3389/fimmu.2016.00472: PMID:27891126
245. Vivier E, Nunès JA, Vély F. Natural killer cell signaling pathways.

- Science. American Association for the Advancement of Science; 2004;306: 1517–9. DOI:10.1126/science.1103478: PMID:15567854
246. Handunnetthi L, Ramagopalan S V, Ebers GC, Knight JC. Regulation of MHC class II gene expression, genetic variation and disease. *Genes Immun.* 2010;11.
247. Rodrigues L, Ferreira P, Ferreira FM, Laugier L, Cabantous S, Cunha Navarro I, et al. Integration of miRNA and gene expression profiles suggest a role for miRNAs in the pathobiological processes of acute *Trypanosoma cruzi* infection. *Sci Rep.* 2017;7: 17990.

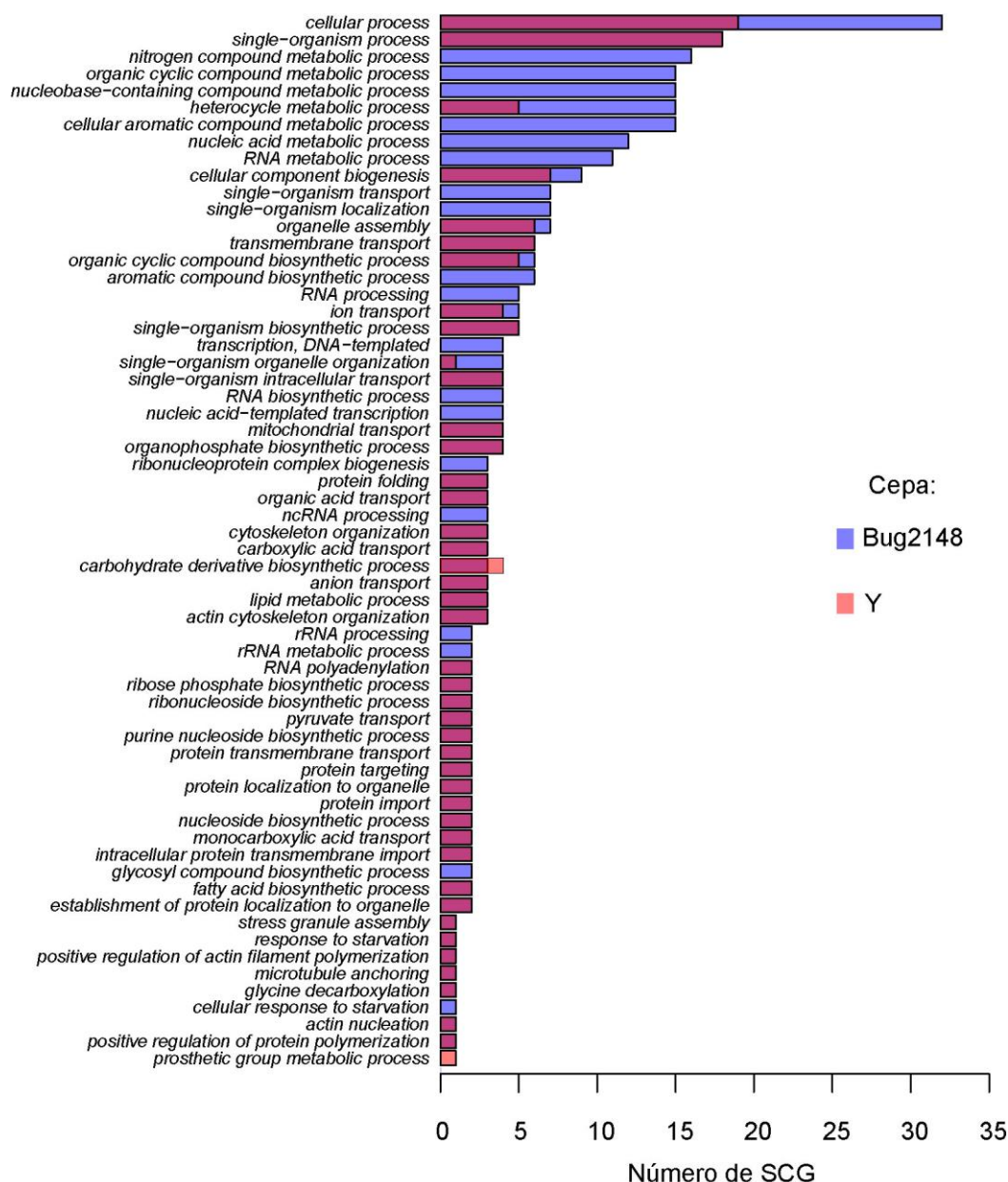




## 7 Anexos

### 7.1 Apéndice A. Material suplementario

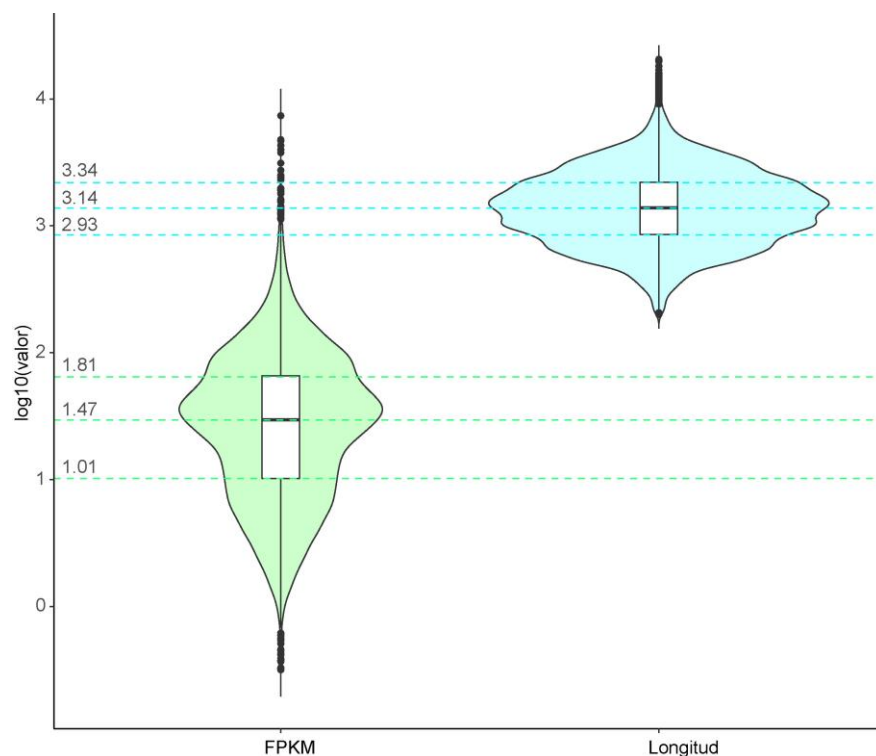
#### 7.1.1 Figuras Suplementarias



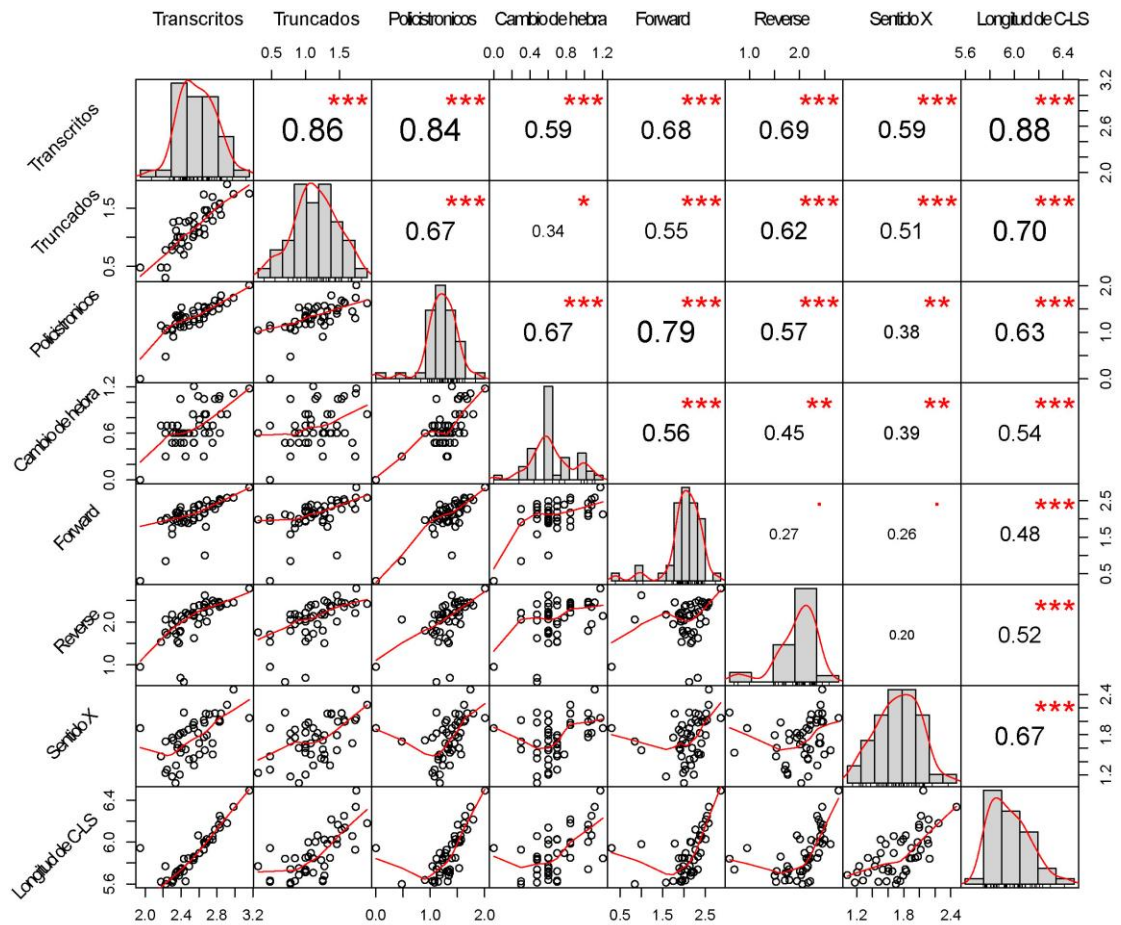
**Figura Suplementaria 1.** Genes de copia única asociados a una función molecular específica por ontología genética a las bases de datos del Tritryp (<http://tritrypdb.org/tritrypdb/>). Redundancia por errores ortográficos en el nombre de las funciones moleculares fue filtrada manualmente en medida de lo posible.



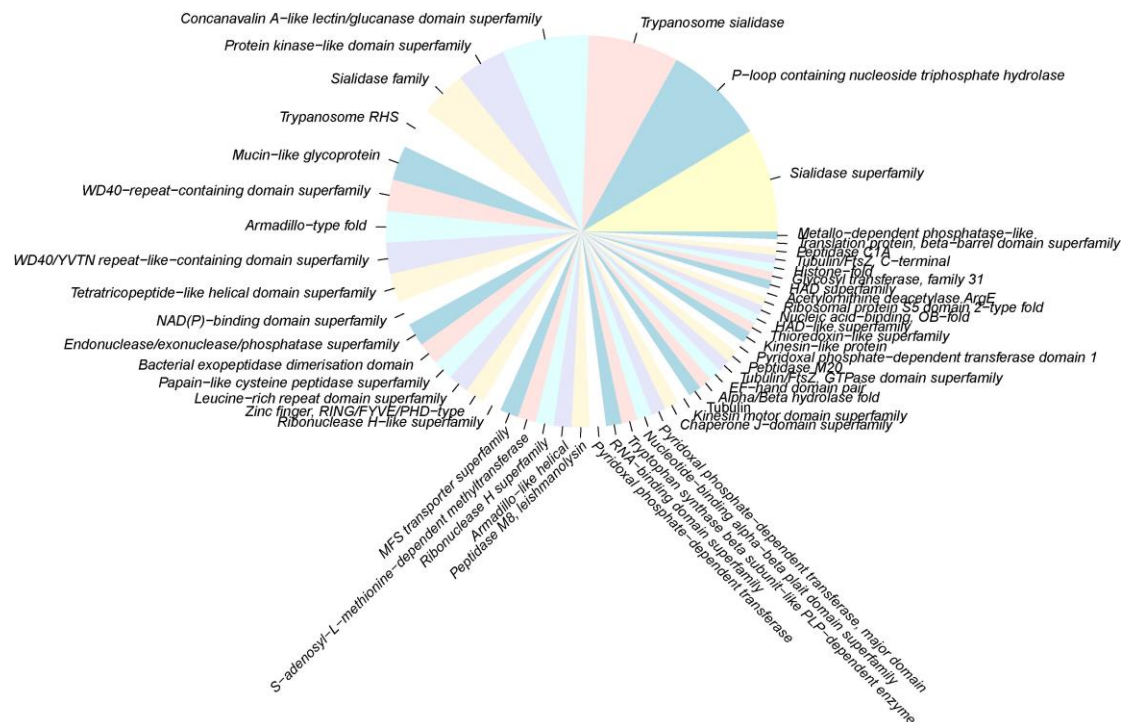
**Figura Suplementaria 2.** Karyoplot de la secuencia anotada codificante de la cepa SylvioX10, antes y después del refinamiento con datos de RNAseq.



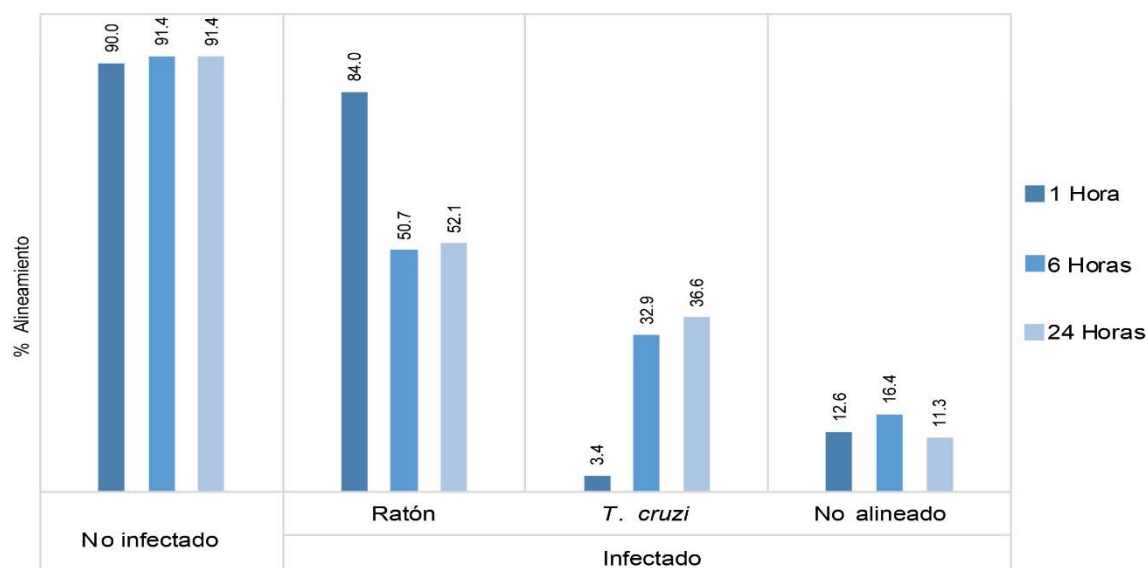
**Figura suplementaria 3.** Abundancia relativa (FPKM) y distribución de longitud del transcriptoma completo de *T. cruzi*. Los cuartiles Q1, Q2 y Q3 se representan en líneas discontinuas.



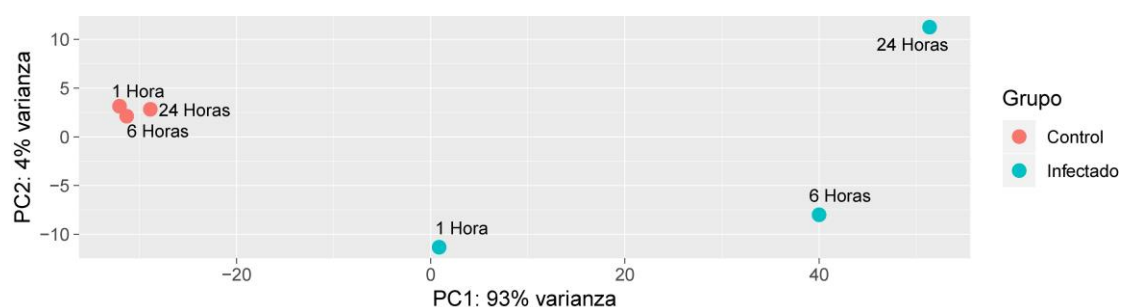
**Figura suplementaria 4.** Correlaciones de spearman entre los datos anotados en el transcriptoma de *Trypanosoma cruzi*. Transcritos se corresponde al total de tr nscritos anotados, policist nicos y truncados se refiere al n mero de estos tr nscritos anotados, el sentido de los tr nscritos Forward, Reverse y ND y SSRs por C-LS. La distribuci n de los datos se muestra en los histogramas y l neas de densidad en diagonal, la matriz inferior contiene las gr ficas de dispersi n bi-variadas y la matriz superior muestra los valores de correlaci n. Los p-valores de las correlaciones se muestran con asteriscos en rojo (0, 0.001, 0.01 se corresponden con “\*\*\*”, “\*\*\*”, “\*”, respectivamente).



**Figura suplementaria 5.** Funciones más abundantes contenidas en el transcriptoma completo de *T. cruzi*. Transcritos de las familias multi-génicas de la forma tripomastigote. La redundancia en los nombres de las funciones por errores ortográficos o de nomenclatura fueron eliminados en medida de lo posible y no se incluyen las funciones hipotéticas.

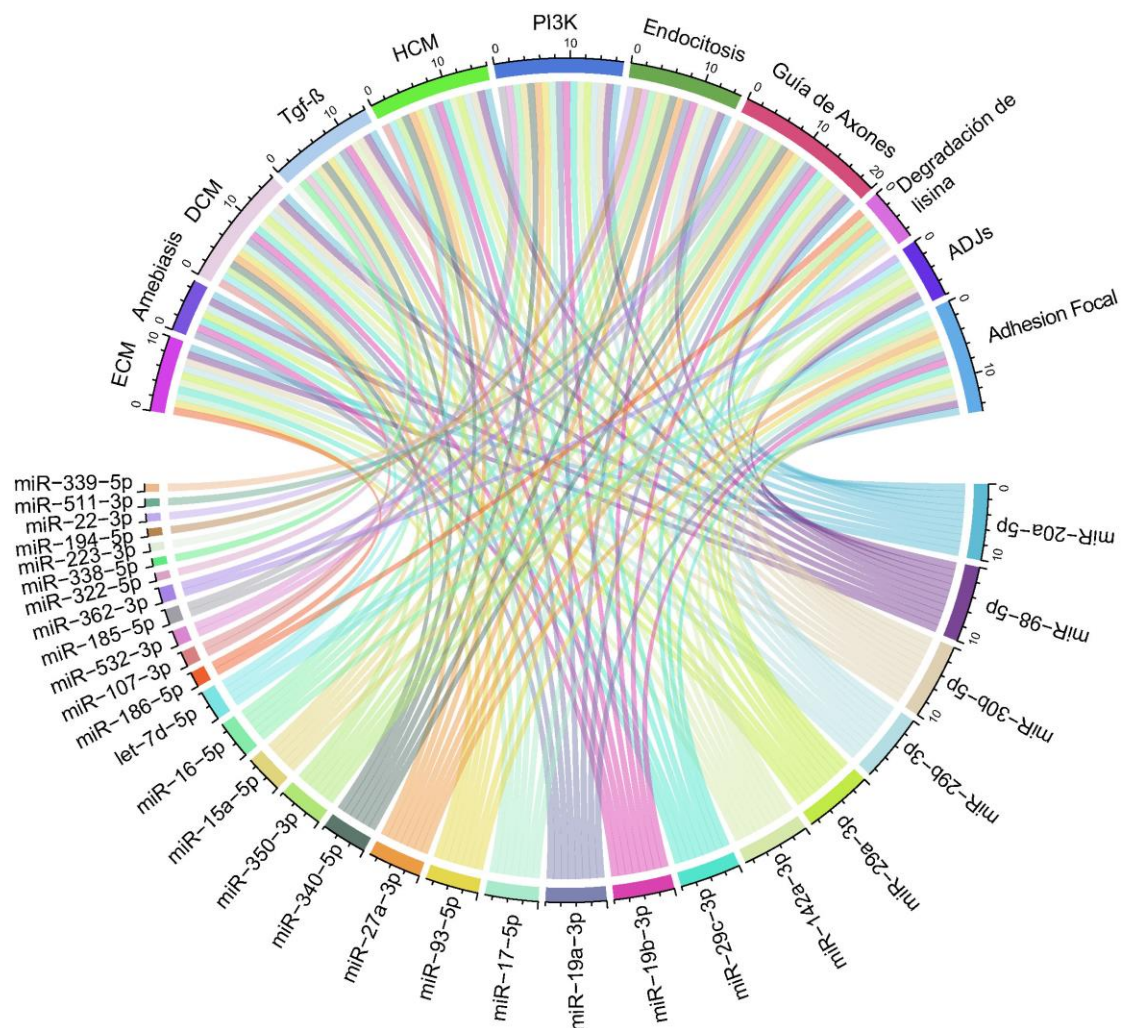


**Figura Suplementaria 6.** Alineamiento de lecturas miARNseq de macrófagos infectados con *T. cruzi* y los controles no infectados. A la izquierda se muestra el porcentaje de lecturas alineadas en las muestras control no infectadas (No infectado) y a la derecha (miARNseq dual) el porcentaje de lecturas alineadas al genoma del huésped, al parásito y no alineadas (Infectado).

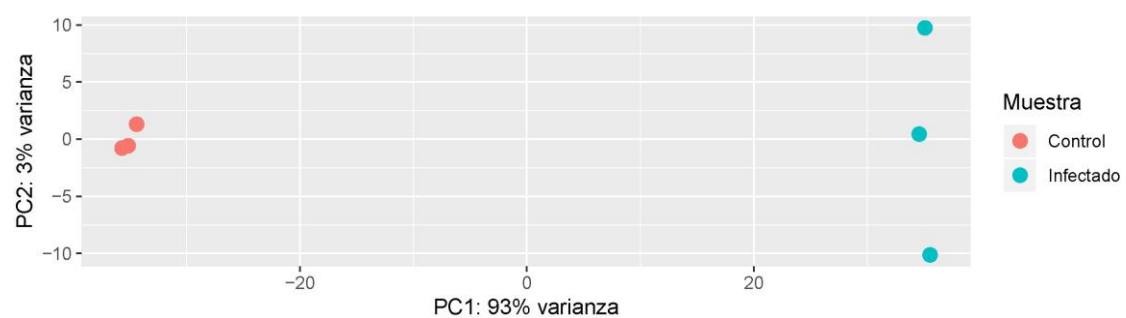


**Figura Suplementaria 7.** Análisis de componentes principales de las lecturas normalizadas de smallRNAseq a 1, 6 y 24 horas post-infección de macrófagos intraperitoneales infectados con *T. cruzi* cepa Y.





**Figura Suplementaria 8.** Rutas funcionalmente enriquecidas ( $p\text{-valor} \leq 0.01$ ) por aumento en las dianas de miARNs inhibidos significativamente a 6 h.p.i, en macrófagos de ratón BALB/c infectados por *T. cruzi*.

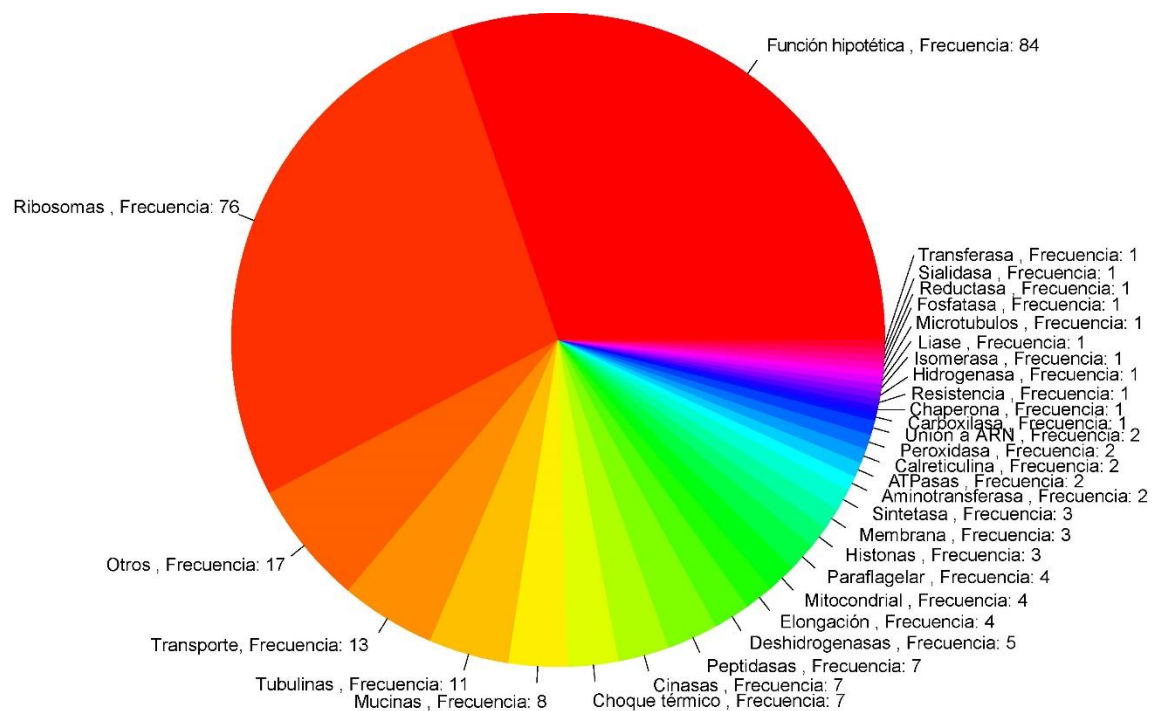


**Figura Suplementaria 9.** Análisis de componentes principales de las lecturas de RNAseq a 24 h.p.i normalizadas con DESeq2.

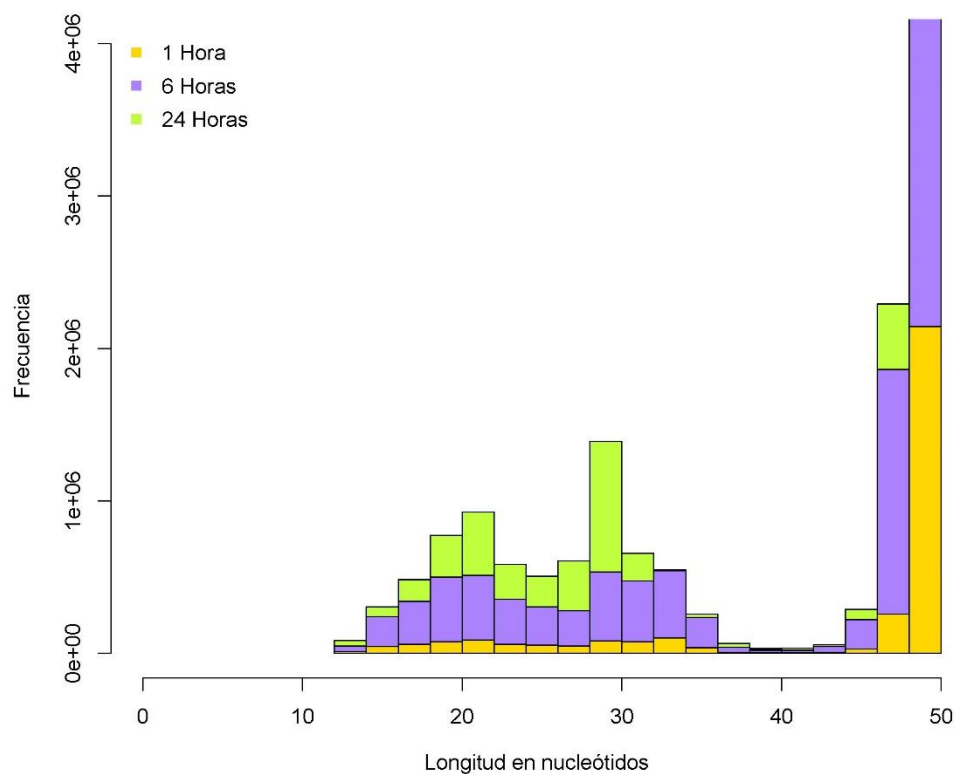


**Figura Suplementaria 10.** Análisis de expresión diferencial a 24 h.p.i de macrófagos de ratón BALB/c infectados con *T. cruzi*. Se remarcen en rojo los genes con expresión diferencial significativa,  $\log_2\text{FoldChange} \leq -3$  (izquierda),  $\geq 3$  (derecha) y  $p\text{-valor-ajustado} \leq 0,001$  (eje Y).

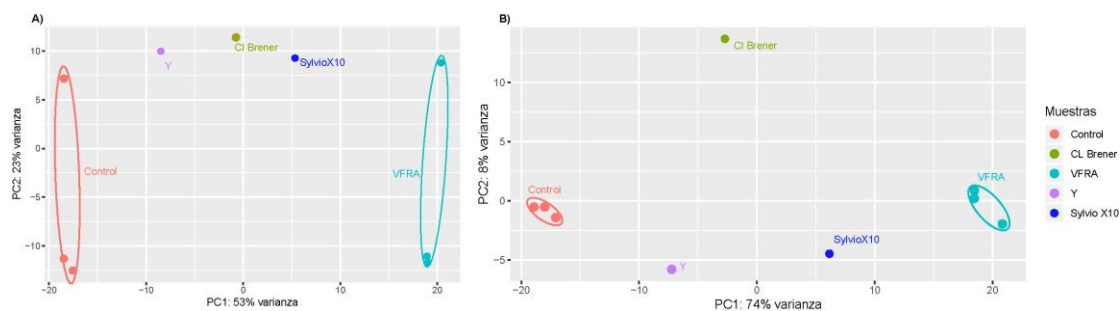




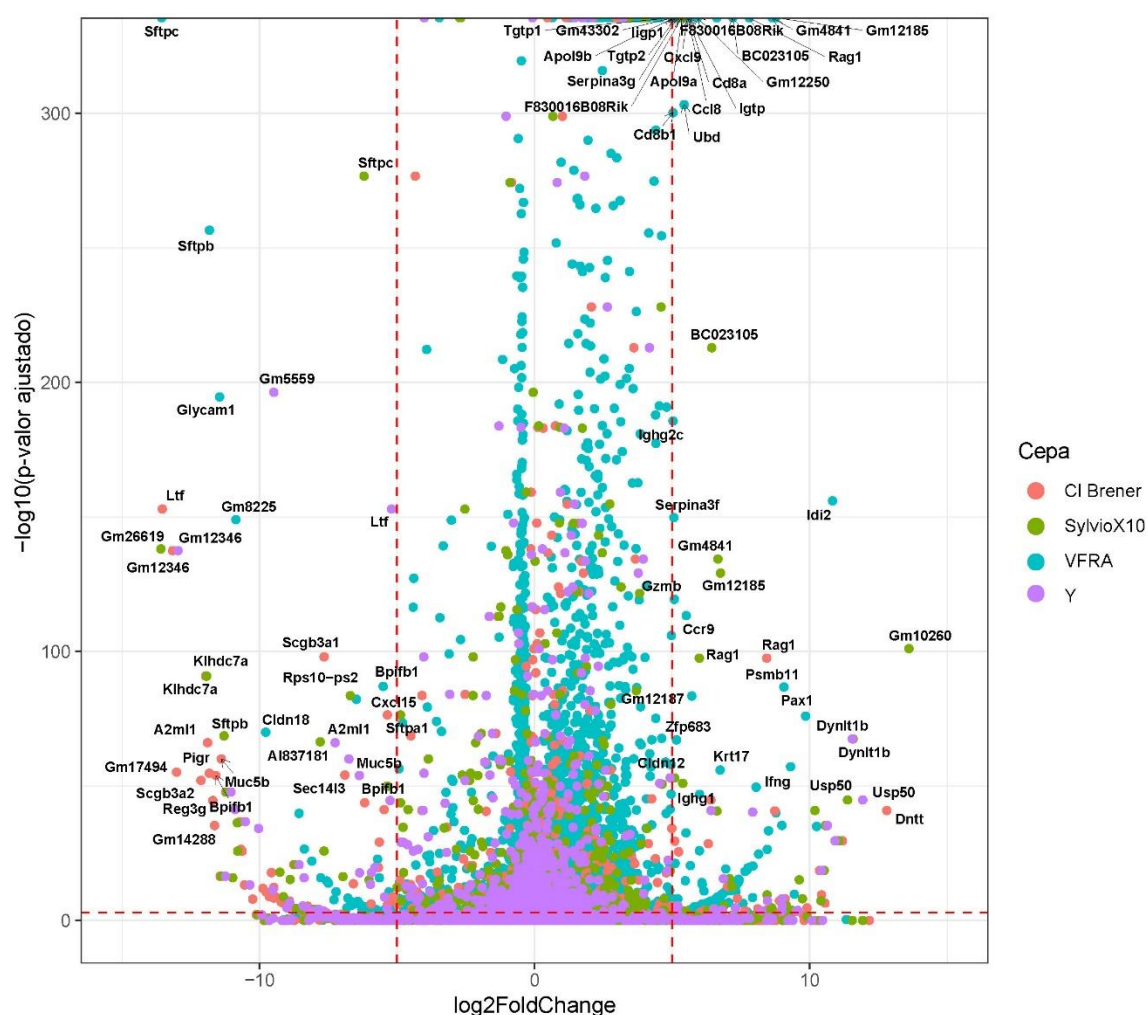
**Figura Suplementaria 11.** Transcritos de *T. cruzi* más abundantes a 24 h.p.i. en macrófagos de ratón BALB/c.



**Figura Suplementaria 12.** Distribución de longitud de las lecturas de secuenciación (miARNseq) correspondientes al genoma del parásito (1, 6 y 24 h.p.i.).



**Figura Suplementaria 13.** Análisis de componentes principales de las muestras de secuenciación de corazón de ratón BALB/c en estado crónico de la enfermedad de Chagas.



**Figura Suplementaria 14.** Análisis de expresión diferencial en corazón de ratón BALB/c en estado crónico de la enfermedad de Chagas versus corazón de ratón no infectado. Se muestran algunos de los genes con expresión diferencial significativa en las diferentes cepas de estudio ( $\log_2\text{Fold} \geq 5$ ,  $\leq -5$  y  $p\text{-valor ajustado} \leq 0,001$ ).

## 7.1.2 Tablas suplementarias

**Tabla Suplementaria 1.** miARNs diferencialmente expresados (sobre-expresados) a 6 h.p.i y sus potenciales ARNm y rutas metabólicas diana.

Gap junction (mmu04540)	Dianas	Proteoglycans in cancer (mmu05205)	Dianas
mmu-miR-543-3p	Drd1a	mmu-miR-182-5p	Cttn
mmu-miR-218-5p	Grm1		Wnt4
	Gucy1a2		Rdx
	Prkg1		Frs2
mmu-miR-1981-5p	Sos1	mmu-miR-495-3p	Vav2
mmu-miR-148a-3p	Sos1		Tfap4
mmu-miR-182-5p	Adcy6		Kras
mmu-miR-381-3p	Grm5		Prkx
mmu-miR-495-3p	Prkx		Wnt5a
	Kras		Cd44
	Map3k2		Igf1
mmu-miR-143-3p	Kras	mmu-miR-1a-3p	Fn1
mmu-miR-183-5p	Gjd2		Frs2
mmu-miR-1a-3p	Gja1	mmu-miR-671-3p	Rps6kb2
Long-term depression (mmu04730)	Dianas	mmu-miR-148a-3p	Cblb
mmu-miR-218-5p	Grm1		Sos1
	Gucy1a2	mmu-miR-143-3p	Kras
	Prkg1	mmu-miR-132-3p	Hbegf
mmu-miR-495-3p	Kras		Rras2
	Gna13	mmu-miR-1981-5p	Sos1
mmu-miR-183-5p	Ppp2ca	mmu-miR-493-5p	Thbs1
	Ppp2cb	mmu-miR-381-3p	Frs2
mmu-miR-1a-3p	Igf1	Thyroid hormone signaling pathway (mmu04919)	Dianas
mmu-miR-677-3p	Gria1	mmu-let-7c-1-3p	Med13
mmu-miR-6539	Gria1	mmu-miR-493-5p	Med13
mmu-miR-143-3p	Kras	mmu-miR-134-5p	Med13
mmu-miR-493-5p	Gria3	mmu-miR-182-5p	Wnt4
			Med1
		mmu-miR-132-3p	Ep300
		mmu-miR-381-3p	Med14
			Ncoa2
		mmu-miR-148a-3p	Med12l
			Med27
		mmu-miR-495-3p	Prkx
			Kras
			Ncoa1
			Crebbp
mmu-miR-382-3p	Rcan2		

	mmu-miR-143-3p	<i>Kras</i>
	mmu-miR-6979-3p	<i>Foxo1</i>

**Tabla Suplementaria 2.** miARNs diferencialmente expresados (inhibidos respecto al control n.i) a 6 h.p.i y sus potenciales ARNm y rutas metabólicas diana.

<b>ECM-receptor interaction</b>	<b>Dianas</b>	<b>TGF-beta signaling pathway</b>	<b>Dianas</b>	<b>Endocytosis</b>	<b>Dianas</b>
mmu-miR-29c-3p	<i>Col2a1</i>	mmu-miR-322-5p	<i>Chrd</i>	mmu-miR-29b-3p	<i>Sh3glb1</i>
	<i>Col3a1</i>	mmu-miR-340-5p	<i>Nog</i>	mmu-miR-98-5p	<i>Adrb3</i>
	<i>Col6a3</i>		<i>Id2</i>		<i>Tgfb1</i>
	<i>Col5a2</i>	mmu-miR-98-5p	<i>Gdf6</i>		<i>Adrb2</i>
	<i>Col4a6</i>		<i>Tgfb1</i>	mmu-miR-340-5p	<i>Arf6</i>
	<i>Col5a3</i>		<i>Acvr1c</i>		<i>Sh3kbp1</i>
	<i>Col4a1</i>	mmu-miR-20a-5p	<i>E2f5</i>		<i>Rab11a</i>
	<i>Col11a1</i>		<i>Zfyve9</i>	mmu-miR-20a-5p	<i>Zfyve9</i>
	<i>Col5a1</i>		<i>Tgfb2</i>		<i>Rab5b</i>
mmu-miR-29b-3p	<i>Col2a1</i>		<i>E2f5</i>		<i>Tgfb2</i>
	<i>Col3a1</i>	mmu-miR-17-5p	<i>E2f5</i>	mmu-miR-19b-3p	<i>Rab11fip5</i>
	<i>Col6a3</i>	mmu-miR-93-5p	<i>E2f5</i>		<i>Zfyve9</i>
	<i>Col5a2</i>	mmu-miR-19b-3p	<i>Zfyve9</i>		<i>Adrb1</i>
	<i>Col4a6</i>		<i>Bambi</i>	mmu-miR-19a-3p	<i>Zfyve9</i>
	<i>Col5a3</i>	mmu-miR-19a-3p	<i>Zfyve9</i>		<i>Vps4b</i>
	<i>Col4a1</i>		<i>Bambi</i>		<i>Adrb1</i>
	<i>Col11a1</i>	mmu-miR-338-5p	<i>Id1</i>	mmu-miR-22-3p	<i>Cyth3</i>
	<i>Col5a1</i>	mmu-miR-142a-3p	<i>Tgfb1</i>		<i>Arrb1</i>
mmu-miR-29a-3p	<i>Col2a1</i>	mmu-miR-30b-5p	<i>Fst</i>	mmu-miR-194-5p	<i>Asap1</i>
	<i>Col3a1</i>	mmu-miR-223-3p	<i>Acvr2a</i>	mmu-miR-142a-3p	<i>Tgfb1</i>
	<i>Col6a3</i>	mmu-miR-16-5p	<i>Acvr2a</i>	mmu-miR-30b-5p	<i>Cblb</i>
	<i>Col5a2</i>		<i>Smad7</i>		<i>Smap1</i>
	<i>Col4a6</i>	mmu-miR-15a-5p	<i>Acvr2a</i>	mmu-miR-350-3p	<i>Eea1</i>
	<i>Col5a3</i>		<i>Smad7</i>	mmu-miR-16-5p	<i>Smad7</i>
	<i>Col4a1</i>	<b>Hypertrophic cardiomyopathy (HCM)</b>	<b>Dianas</b>	mmu-miR-15a-5p	<i>Smad7</i>
	<i>Col5a1</i>	mmu-miR-350-3p	<i>Cacnb4</i>	mmu-miR-27a-3p	<i>Argef1</i>
	<i>Col3a1</i>		<i>Dmd</i>	mmu-miR-532-3p	<i>Fgfr2</i>
mmu-miR-98-5p	<i>Col27a1</i>	mmu-miR-340-5p	<i>Actc1</i>	<b>Axon guidance</b>	<b>Dianas</b>
	<i>Col1a2</i>		<i>Dmd</i>		<i>Sema7a</i>
					<i>Epha4</i>
mmu-miR-186-5p	<i>Sv2c</i>	mmu-miR-107-3p	<i>Cacna2d1</i>	mmu-miR-20a-5p	<i>Ppp3r1</i>
mmu-miR-20a-5p	<i>Itgb8</i>	mmu-miR-30b-5p	<i>Actc1</i>		<i>Epha7</i>
mmu-miR-17-5p	<i>Itgb8</i>		<i>Itga8</i>		<i>Epha7</i>
mmu-miR-93-5p	<i>Itgb8</i>		<i>Sgcb</i>	mmu-miR-16-5p	<i>Sema6d</i>
mmu-miR-142a-3p	<i>Itgav</i>	mmu-let-7d-5p	<i>Atp2a2</i>		<i>Unc5c</i>
mmu-miR-30b-5p	<i>Itga8</i>		<i>Actg1</i>	mmu-miR-339-5p	<i>Rac1</i>
		mmu-miR-27a-3p	<i>Ttn</i>	mmu-miR-142a-3p	

<b>Amoebiasis</b>	<b>Dianas</b>		<i>Cacna2d3</i>		<i>Rock2</i>
mmu-miR-98-5p	<i>Col3a1</i>	mmu-miR-29c-3p	<i>Igf1</i>	mmu-miR-27a-3p	<i>Sema6a</i>
	<i>Col27a1</i>		<i>Tpm1</i>		<i>Sema7a</i>
	<i>Col1a2</i>	mmu-miR-29b-3p	<i>Igf1</i>	mmu-miR-340-5p	<i>Nfatc3</i>
mmu-miR-29c-3p	<i>Col2a1</i>		<i>Tpm1</i>	mmu-miR-30b-5p	<i>Nfatc3</i>
	<i>Col3a1</i>	mmu-miR-29a-3p	<i>Igf1</i>		<i>Rasa1</i>
	<i>Col5a2</i>		<i>Tpm1</i>		<i>Sema3a</i>
	<i>Col4a6</i>	mmu-miR-142a-3p	<i>Itgav</i>	mmu-miR-511-3p	<i>Sema3a</i>
	<i>Col5a3</i>	mmu-miR-19b-3p	<i>Cacna1c</i>	mmu-miR-17-5p	<i>Epha4</i>
	<i>Col4a1</i>	mmu-miR-19a-3p	<i>Cacna1c</i>		<i>Ppp3r1</i>
	<i>Col11a1</i>	mmu-miR-98-5p	<i>Actg1</i>	mmu-miR-93-5p	<i>Epha4</i>
	<i>Col5a1</i>	mmu-miR-20a-5p	<i>Itgb8</i>		<i>Ppp3r1</i>
mmu-miR-29b-3p	<i>Col2a1</i>	mmu-miR-17-5p	<i>Itgb8</i>	mmu-miR-15a-5p	<i>Sema6d</i>
	<i>Col3a1</i>	mmu-miR-93-5p	<i>Itgb8</i>	mmu-miR-98-5p	<i>Sema4f</i>
	<i>Col5a2</i>	<b>PI3K-Akt signaling pathway</b>	<b>Dianas</b>		<i>Plxnc1</i>
	<i>Col4a6</i>		<i>Col2a1</i>	mmu-miR-29c-3p	<i>Robo1</i>
	<i>Col5a3</i>		<i>Col3a1</i>	mmu-miR-29b-3p	<i>Robo1</i>
	<i>Col4a1</i>		<i>Col6a3</i>	mmu-miR-29a-3p	<i>Robo1</i>
	<i>Col11a1</i>		<i>Igf1</i>	mmu-miR-350-3p	<i>Ppp3r1</i>
	<i>Col5a1</i>		<i>Vegfa</i>	mmu-miR-19b-3p	<i>Plxnc1</i>
mmu-miR-29a-3p	<i>Col2a1</i>	mmu-miR-29c-3p	<i>Col5a2</i>	mmu-miR-19a-3p	<i>Plxnc1</i>
	<i>Col3a1</i>		<i>Col4a6</i>	mmu-miR-362-3p	<i>Nfatc4</i>
	<i>Col5a2</i>		<i>Col5a3</i>	mmu-miR-185-5p	<i>Ephb2</i>
	<i>Col4a6</i>		<i>Col4a1</i>	<b>Lysine degradation</b>	<b>Dianas</b>
	<i>Col5a3</i>		<i>Col11a1</i>	mmu-miR-27a-3p	<i>Kmt2c</i>
	<i>Col4a1</i>		<i>Pten</i>		<i>Setd2</i>
	<i>Col5a1</i>		<i>Col5a1</i>		<i>Colgalt2</i>
mmu-miR-19b-3p	<i>Pik3r3</i>		<i>Col2a1</i>	mmu-miR-30b-5p	<i>Kmt2c</i>
	<i>Adcy1</i>		<i>Col3a1</i>		<i>Setd7</i>
mmu-miR-20a-5p	<i>Rab5b</i>		<i>Col6a3</i>	mmu-miR-29c-3p	<i>Nsd1</i>
mmu-miR-19a-3p	<i>Pik3r3</i>		<i>Igf1</i>	mmu-miR-186-5p	<i>Nsd1</i>
			<i>Vegfa</i>	mmu-miR-29b-3p	<i>Nsd1</i>
			<i>Col5a2</i>	mmu-miR-29a-3p	<i>Nsd1</i>
<b>Dilated cardiomyopathy</b>	<b>Dianas</b>	mmu-miR-29b-3p	<i>Col4a6</i>	mmu-miR-350-3p	<i>Dot1l</i>
mmu-miR-30b-5p	<i>Actc1</i>		<i>Col5a3</i>	<b>Adherens junction</b>	<b>Dianas</b>
	<i>Itga8</i>		<i>Col4a1</i>	mmu-miR-362-3p	<i>Pvrl4</i>
	<i>Sgcb</i>		<i>Col11a1</i>	mmu-miR-350-3p	<i>Tcf7l2</i>
mmu-let-7d-5p	<i>Atp2a2</i>		<i>Pten</i>	mmu-miR-142a-3p	<i>Wasl</i>
	<i>Actg1</i>		<i>Col5a1</i>		<i>Tgfb1</i>
mmu-miR-340-5p	<i>Actc1</i>	mmu-miR-29a-3p	<i>Col2a1</i>	mmu-miR-98-5p	<i>Rac1</i>
	<i>Dmd</i>		<i>Col3a1</i>		<i>Fer</i>
mmu-miR-350-3p	<i>Cacnb4</i>		<i>Col6a3</i>	mmu-let-7d-5p	<i>Actg1</i>
	<i>Dmd</i>		<i>Igf1</i>	mmu-miR-98-5p	<i>Tgfb1</i>
mmu-miR-98-5p	<i>Actg1</i>		<i>Vegfa</i>		<i>Actg1</i>
mmu-miR-19b-3p	<i>Adcy1</i>				

	<i>Adrb1</i>		<i>Col5a2</i>		<i>Pvrl3</i>
	<i>Cacna1c</i>		<i>Col4a6</i>	mmu-miR-20a-5p	<i>Tgfb2</i>
mmu-miR-19a-3p	<i>Adrb1</i>		<i>Col5a3</i>	mmu-miR-30b-5p	<i>Snai1</i>
	<i>Cacna1c</i>		<i>Col4a1</i>	<b>Focal adhesion</b>	<b>Dianas</b>
mmu-miR-27a-3p	<i>Ttn</i>		<i>Pten</i>	mmu-miR-30b-5p	<i>Ppp1r12a</i>
	<i>Cacna2d3</i>		<i>Col5a1</i>		<i>Itga8</i>
mmu-miR-29c-3p	<i>Igf1</i>	mmu-miR-98-5p	<i>Col3a1</i>	mmu-miR-16-5p	<i>Vegfa</i>
	<i>Tpm1</i>		<i>Ngf</i>		<i>Col2a1</i>
mmu-miR-29b-3p	<i>Igf1</i>		<i>Col27a1</i>		<i>Col3a1</i>
	<i>Tpm1</i>		<i>Col1a2</i>		<i>Col6a3</i>
mmu-miR-29a-3p	<i>Igf1</i>		<i>Tsc1</i>		<i>Igf1</i>
	<i>Tpm1</i>		<i>Lpar6</i>		<i>Vegfa</i>
mmu-miR-20a-5p	<i>Itgb8</i>	mmu-miR-27a-3p	<i>Creb1</i>	mmu-miR-29c-3p	<i>Col5a2</i>
mmu-miR-17-5p	<i>Itgb8</i>		<i>Grb2</i>		<i>Col4a6</i>
mmu-miR-93-5p	<i>Itgb8</i>		<i>Csf1</i>		<i>Col5a3</i>
mmu-miR-107-3p	<i>Cacna2d1</i>	mmu-miR-142a-3p	<i>Prlr</i>		<i>Col4a1</i>
mmu-miR-142a-3p	<i>Itgav</i>		<i>Rac1</i>		<i>Col11a1</i>
			<i>Itgav</i>		<i>Pten</i>
		mmu-miR-532-3p	<i>Fgfr2</i>		<i>Col5a1</i>
		mmu-miR-20a-5p	<i>Rbl2</i>		<i>Col2a1</i>
			<i>Mcl1</i>		<i>Col3a1</i>
			<i>Itgb8</i>		<i>Col6a3</i>
		mmu-miR-17-5p	<i>Itgb8</i>		<i>Igf1</i>
		mmu-miR-93-5p	<i>Itgb8</i>		<i>Vegfa</i>
		mmu-miR-19b-3p	<i>Pik3r3</i>	mmu-miR-29b-3p	<i>Col5a2</i>
			<i>Atf2</i>		<i>Col4a6</i>
		mmu-miR-19a-3p	<i>Pik3r3</i>		<i>Col5a3</i>
			<i>Atf2</i>		<i>Col4a1</i>
		mmu-miR-15a-5p	<i>Vegfa</i>		<i>Col11a1</i>
			<i>Ghr</i>		<i>Pten</i>
			<i>Ccne1</i>		<i>Col5a1</i>
			<i>Bcl2</i>		<i>Col2a1</i>
			<i>Sgk1</i>		<i>Col3a1</i>
			<i>Myb</i>		<i>Col6a3</i>
		mmu-miR-340-5p	<i>Creb5</i>		<i>Igf1</i>
		mmu-miR-16-5p	<i>Vegfa</i>	mmu-miR-29a-3p	<i>Vegfa</i>
			<i>Ghr</i>		<i>Col5a2</i>
			<i>Ccne1</i>		<i>Col4a6</i>
			<i>Sgk1</i>		<i>Col5a3</i>
			<i>Myb</i>		<i>Col4a1</i>
		mmu-miR-30b-5p	<i>Itga8</i>		<i>Pten</i>
		mmu-miR-185-5p	<i>Phlpp2</i>		<i>Col5a1</i>
				mmu-miR-15a-5p	<i>Vegfa</i>
					<i>Bcl2</i>

		mmu-miR-19b-3p	<i>Pik3r3</i>
		mmu-miR-19a-3p	<i>Pik3r3</i>
		mmu-miR-142a-3p	<i>Rac1</i>
			<i>Rock2</i>
			<i>Itgav</i>
		mmu-miR-20a-5p	<i>Itgb8</i>
		mmu-miR-17-5p	<i>Itgb8</i>
		mmu-miR-93-5p	<i>Itgb8</i>
		mmu-let-7d-5p	<i>Actg1</i>
		mmu-miR-98-5p	<i>Col3a1</i>
			<i>Col27a1</i>
			<i>Col1a2</i>
			<i>Actg1</i>
		mmu-miR-27a-3p	<i>Grb2</i>
			<i>Ppp1cc</i>

**Tabla Suplementaria 3.** Enriquecimiento funcional a 24 h.p.i de macrófagos intraperitoneales de ratón BALB/c infectados con *T. cruzi*.

Ruta	Enriquecimiento Normalizado	FDR	Genes	Entrada
<i>TNF pathway</i> <i>signaling</i>	2.97561557	4.15E-06	52	Birc3, Ccl5, Cebpb, Cxcl1, Cxcl2, Cxcl3, Icam1, Ifi47, Il1b, Il6, Jag1, Junb, Lif, Mmp14, Mmp9, Nfkbia, Ptgs2, Socs3, Tnf, Tnfrsf1b, Traf1, Traf2
<i>IL-17 pathway</i> <i>signaling</i>	2.9270854	4.15E-06	32	Ccl7, Cebpb, Cxcl1, Cxcl2, Cxcl3, Cxcl5, Il17ra, Il1b, Il6, Lcn2, Mmp9, Nfkbia, Ptgs2, S100a8, Tnf, Traf2
<i>Ribosome</i>	2.75585833	4.15E-06	112	Fau, Mrpl1, Mrpl11, Mrpl12, Mrpl13, Mrpl15, Mrpl2, Mrpl20, Mrpl21, Mrpl22, Mrpl23, Mrpl24, Mrpl27, Mrpl28, Mrpl30, Mrpl32, Mrpl34, Mrpl35, Mrpl36, Mrpl9, Mrps10, Mrps14, Mrps16, Mrps18a, Mrps18c, Mrps7, Mrps9, Rpl11, Rpl12, Rpl13, Rpl13a, Rpl14, Rpl15, Rpl17, Rpl18, Rpl18a, Rpl19, Rpl22, Rpl22l1, Rpl23, Rpl23a, Rpl24, Rpl26, Rpl27, Rpl27a, Rpl3, Rpl30, Rpl31, Rpl32, Rpl34, Rpl34-ps1, Rpl35, Rpl35a, Rpl36, Rpl36a, Rpl36al, Rpl37, Rpl37a, Rpl37rt, Rpl38, Rpl39, Rpl41, Rpl5, Rpl6, Rpl7, Rpl7a, Rpl8, Rpl9, Rplp0, Rplp1, Rplp2, Rps10, Rps12, Rps13, Rps14, Rps15, Rps15a, Rps16, Rps17, Rps18, Rps19, Rps2, Rps20, Rps21, Rps23, Rps24, Rps25, Rps26, Rps27, Rps27a, Rps27l, Rps29, Rps3, Rps3a1, Rps5, Rps6, Rps7, Rps8, Rps9, Rpsa, Rsl24d1, Uba52, mt-Rnr1, mt-Rnr2
<i>NOD-like receptor signaling pathway</i>	2.72540769	4.15E-06	80	Birc3, Casp4, Ccl5, Cxcl1, Cxcl2, Cxcl3, Cybb, Gbp2, Gbp2b, Gbp3, Gbp7, Ifi204, Il1b, Il6, Irf7, Irf9, Nampt, Nfkbia, Nfkibib, Nlrp3, Oas1a, Oas1g, Oas2, Oas3, Tank, Tnf, Traf2, Vdac3

Ruta	Enriquecimiento Normalizado	FDR	Genes	Entrada
<i>Herpes simplex infection</i>	2.62895023	4.15E-06	86	Ccl5, Ddx58, H2-K1, H2-M2, H2-Q6, Ifih1, Ifit1, Il1b, Il6, Irf7, Irf9, Jak2, Nfkb1a, Nfkb1b, Oas1a, Oas1g, Oas2, Oas3, Pirla, Socs3, Tap1, Tlr2, Tnf, Traf1, Traf2
<i>Legionellosis</i>	2.6195363	4.15E-06	29	Cd14, Cxcl1, Cxcl2, Cxcl3, Il1b, Il6, Nfkb2, Nfkb1a, Tlr2, Tnf
<i>Cytokine-cytokine receptor interaction</i>	2.58877534	4.15E-06	59	Ccl3, Ccl5, Ccl7, Ccl9, Cxcl1, Cxcl16, Cxcl2, Cxcl3, Cxcl5, Il17ra, Il1a, Il1b, Il33, Il6, Inhba, Lif, Osm, Pf4, Tnf, Tnfrsf12a, Tnfrsf1b
<i>Cytosolic DNA-sensing pathway</i>	2.58065439	4.15E-06	29	Ccl4, Ccl5, Ddx58, Ifi202b, Il1b, Il33, Il6, Irf7, Nfkb1a, Nfkb1b, Polr2h, Trex1, Zbp1
<i>Influenza A</i>	2.48790235	4.15E-06	73	Ccl5, Ddx58, Icam1, Ifih1, Il1a, Il1b, Il33, Il6, Irf7, Irf9, Jak2, Nfkb1a, Nfkb1b, Nlrp3, Oas1a, Oas1g, Oas2, Oas3, Rsad2, Socs3, Tlr7, Tnf
<i>NF-kappa B signaling pathway</i>	2.47852702	4.15E-06	50	Bcl2a1a, Bcl2a1b, Bcl2a1d, Birc3, Ccl4, Cd14, Cxcl2, Ddx58, Icam1, Il1b, Lbp, Malt1, Nfkb2, Nfkb1a, Ptgs2, Relb, Tnf, Traf1, Traf2
<i>Proteasome</i>	2.45630946	4.15E-06	39	Adrm1, Pomp, Psma1, Psma2, Psma3, Psma4, Psma5, Psma6, Psma7, Psmb1, Psmb10, Psmb2, Psmb3, Psmb4, Psmb5, Psmb6, Psmb7, Psmb8, Psmb9, Psmc2, Psmc3, Psmc4, Psmc6, Psmd1, Psmd11, Psmd13, Psmd14, Psmd4, Psmd6, Psmd8, Psme1, Psme2, Psme2b, Sem1
<i>Parkinson disease</i>	2.45215462	4.15E-06	80	Adora2a, Atp5c1, Atp5e, Atp5f1, Atp5g2, Atp5h, Atp5j, Atp5o, Cox4i1, Cox5b, Cox6a1, Cox6b1, Cox6c, Cox7a2, Cox7a2l, Cox7b, Cox7c, Cox8a, Cyc1, Cycs, Lrrk2, Ndufa1, Ndufa11, Ndufa12, Ndufa13, Ndufa2, Ndufa3, Ndufa4, Ndufa5, Ndufa6, Ndufa7, Ndufa9, Ndufab1, Ndufb11, Ndufb2, Ndufb3, Ndufb4, Ndufb5, Ndufb6, Ndufb7, Ndufb8, Ndufb9, Ndufc1, Ndufc2, Ndufs3, Ndufs4, Ndufs5, Ndufs6, Ndufs7, Ndufv1, Ndufv2, Sdhb, Uba7, Ubb, Ube2j2, Uqcr10, Uqcr11, Uqcrb, Uqcrc2, Uqcrcs1, Uqcrcq, Vdac3
<i>Rheumatoid arthritis</i>	2.40753131	4.15E-06	38	Acp5, Ccl3, Ccl5, Cxcl5, Icam1, Il1a, Il1b, Il6, Itgal, Tlr2, Tnf, Vegfa
<i>Salmonella infection</i>	2.37489397	4.15E-06	38	Ccl3, Cd14, Cxcl1, Cxcl2, Cxcl3, Il1a, Il1b, Il6, Lbp
<i>Leishmaniasis</i>	2.36990109	4.15E-06	32	Cybb, Il1a, Il1b, Jak2, Marcks1, Nfkb1a, Nfkb1b, Ptgs2, Tlr2, Tnf
<i>Inflammatory bowel disease (IBD)</i>	2.34813238	4.15E-06	17	Il1a, Il1b, Il6, Tlr2, Tnf
<i>Steroid biosynthesis</i>	-2.32281017	4.15E-06	13	Cyp51, Dhcr24, Dhcr7, Fdft1, Hsd17b7, Lipa, Lss, Msmo1, Nsdhl, Sc5d, Soat1, Sqle
<i>Non-alcoholic fatty liver disease (NAFLD)</i>	2.2963261	4.15E-05	94	Atf4, Bax, Bid, Casp7, Cox4i1, Cox5b, Cox6c, Cox7a2, Cox7a2l, Cox7b, Cox7c, Cox8a, Cycs, Ddit3, Il1a, Il1b, Il6, Ndufa1, Ndufa11, Ndufa12, Ndufa13, Ndufa2, Ndufa3, Ndufa4, Ndufa6, Ndufa7, Ndufab1, Ndufb11, Ndufb2, Ndufb3, Ndufb4, Ndufb5, Ndufb6, Ndufb8, Ndufb9, Ndufc1, Ndufc2, Ndufs4, Ndufs5, Ndufs6, Ndufv2, Nfkb1, Sdhb, Socs3, Tnf, Traf2, Uqcr10, Uqcr11, Uqcrb, Uqcrcs1, Uqcrcq



Ruta	Enriquecimiento Normalizado	FDR	Genes	Entrada
<i>Type 1 diabetes mellitus</i>	2.29765298	4.36E-05	17	Gm11127, H2-K1, H2-M2, H2-M3, H2-Q1, H2-Q4, H2-Q6, H2-T10, H2-T22, H2-T23, Il1a, Il1b, Tnf
<i>Measles</i>	2.30099368	4.59E-05	57	Ddx58, Ifih1, Il1a, Il1b, Il6, Irf7, Irf9, Jak2, Nfkbia, Nfkbib, Oas1a, Oas1g, Oas2, Oas3, Tlr2, Tlr7
<i>Toll-like receptor signaling pathway</i>	2.3361216	4.85E-05	42	Ccl3, Ccl4, Ccl5, Cd14, Ifnar2, Il1b, Il6, Irf7, Lbp, Map2k1, Myd88, Nfkbia, Stat1, Ticam2, Tlr2, Tlr7, Tnf
<i>Graft-versus-host disease</i>	2.34465739	5.13E-05	17	Gm11127, H2-K1, H2-M2, H2-M3, H2-Q1, H2-Q4, H2-Q6, H2-T10, H2-T22, H2-T23, Il1a, Il1b, Il6, Tnf
<i>Malaria</i>	2.2594155	2.38E-04	17	Icam1, Il1b, Il6, Itgal, Tlr2, Tnf
<i>C-type lectin receptor signaling pathway</i>	2.25075372	2.66E-04	53	Bcl3, Clec4d, Clec4e, Clec4n, Il1b, Il6, Irf9, Malt1, Mapkapk2, Nfkb2, Nfkbia, Nlrp3, Plk3, Ptgs2, Relb, Rras, Stat1, Tnf
<i>Chagas disease (American trypanosomiasis)</i>	2.24012393	2.91E-04	44	Ccl3, Ccl5, Il1b, Il6, Nfkbia, Serpine1, Tlr2, Tnf
<i>Oxidative phosphorylation</i>	2.22624106	3.84E-04	87	Atp5e, Atp5f1, Atp5h, Atp5j, Atp5k, Atp5l, Atp5o, Atp6v0b, Atp6v0e, Atp6v1e1, Atp6v1f, Atp6v1g1, Atp6v1h, Cox17, Cox4i1, Cox5b, Cox6c, Cox7a2, Cox7a2l, Cox7b, Cox7c, Cox8a, Ndufa1, Ndufa11, Ndufa12, Ndufa13, Ndufa2, Ndufa3, Ndufa4, Ndufa5, Ndufa6, Ndufa7, Ndufab1, Ndufb11, Ndufb2, Ndufb3, Ndufb4, Ndufb5, Ndufb6, Ndufb7, Ndufb8, Ndufb9, Ndufc1, Ndufc2, Ndufs3, Ndufs4, Ndufs5, Ndufs6, Ndufs7, Ndufv1, Ndufv2, Ppa1, Sdhb, Uqcr10, Uqcr11, Uqcrb, Uqcrrs1, Uqcrrq
<i>Biosynthesis of amino acids</i>	2.18710077	7.38E-04	34	Aldh18a1, Arg1, Arg2, Asns, Ass1, Bcat1, Cth, Eno1, Eno1b, Gpt2, Pfkf, Pgk1, Phgdh, Psph, Rpe, Shmt2, Tpi1
<i>Epstein-Barr virus infection</i>	2.14752065	0.00109874	112	Bax, Cxcs, Ddx58, Gadd45a, H2-K1, H2-M2, H2-Q1, H2-Q4, H2-Q6, H2-T10, H2-T22, Icam1, Ifnar2, Il6, Irf7, Irf9, Isg15, Itgal, Myd88, Nfkb2, Nfkbia, Nfkbib, Nfkbie, Oas1a, Oas1g, Oas2, Oas3, Psmc6, Psmc8, Relb, Sem1, Stat1, Tap1, Tlr2, Tnf, Traf2

**Tabla Suplementaria 4.** Rutas enriquecidas funcionalmente tras el análisis GSEA del transcriptoma completo de corazón de ratón BALB/c en estado crónico de la enfermedad de Chagas.

Cepas	Total	Rutas comunes/únicas
Cl Brener, VFRA, Sylvio X10	1	<i>T cell receptor signaling pathway</i>
VFRA, Sylvio X10, Y	14	<i>Viral myocarditis</i>
		<i>Th17 cell differentiation</i>
		<i>Inflammatory bowel disease (IBD)</i>
		<i>Antigen processing and presentation</i>
		<i>Type I diabetes mellitus</i>
		<i>Human T-cell leukemia virus 1 infection</i>
		<i>Th1 and Th2 cell differentiation</i>
		<i>Graft-versus-host disease</i>
		<i>Autoimmune thyroid disease</i>
		<i>Natural killer cell mediated cytotoxicity</i>
		<i>Human immunodeficiency virus 1 infection</i>
		<i>Epstein-Barr virus infection</i>
		<i>Herpes simplex infection</i>
		<i>Allograft rejection</i>
VFRA, Sylvio X10	7	<i>Hematopoietic cell lineage</i>
		<i>Chagas disease (American trypanosomiasis)</i>
		<i>Tuberculosis</i>
		<i>Cellular senescence</i>
		<i>Primary immunodeficiency</i>
		<i>Toxoplasmosis</i>
		<i>Cytokine-cytokine receptor interaction</i>
Sylvio X10, Y	1	<i>Viral carcinogenesis</i>
VFRA, Y	1	<i>Leishmaniasis</i>
Sylvio X10	5	<i>NOD-like receptor signaling pathway</i>
		<i>Kaposi sarcoma-associated herpesvirus infection</i>
		<i>Influenza A</i>
		<i>Endocytosis</i>
		<i>Human cytomegalovirus infection</i>
VFRA	1	<i>Rheumatoid arthritis</i>

**Tabla suplementaria 5.** Genes comunes diferencial y significativamente sobre-expresados entre rutas relacionadas con enfermedades autoinmunes y citotoxicidad. Análisis de expresión diferencial en corazón de ratón BALB/c en estado crónico de la enfermedad de Chagas.

Rutas	Genes
Rechazo de aloinjerto, Artritis reumatoide, Enfermedad autoinmune tiroidea, Enfermedad Injerto-contrahuésped	<i>H2-DMb1, H2-Aa, H2-Ab1, H2-DMb2, Cd86, H2-DMa, H2-Oa, H2-Eb1, H2-Ea-ps</i>
Rechazo de aloinjerto, Enfermedad tiroidea autoinmune, Enfermedad injerto-contrahuésped, Citotoxicidad mediada por células NK	<i>H2-K1, FasI, Prf1, H2-T23, H2-D1, Gzmb</i>
Rechazo de aloinjerto, Artritis reumatoide, Enfermedad injerto-contrahuésped, Citotoxicidad mediada por células NK	<i>Tnf</i>
Rechazo de aloinjerto, Enfermedad autoinmune tiroidea, Enfermedad injerto-contrahuésped	<i>H2-Q2, H2-Q6, H2-T3, H2-BI, H2-Q4, H2-Q1, H2-M2, H2-Q7, H2-T22, Il2, H2-T-ps, H2-M3</i>
Rechazo de aloinjerto, Artritis reumatoide, Enfermedad injerto-contrahuésped	<i>Ifng, Cd80, H2-Ob</i>
Rechazo de aloinjerto, Enfermedad injerto-contrahuésped	<i>H2-T24</i>
Artritis reumatoide, Enfermedad injerto-contrahuésped	<i>Il6, Il1b</i>
Enfermedad injerto-contrahuésped, Citotoxicidad mediada por células NK	<i>Klrd1, Klrc1,, Klra1, Klra3</i>
Rechazo de aloinjerto, Enfermedad autoinmune tiroidea	<i>Cd40lg</i>
Artritis reumatoide, Citotoxicidad mediada por células NK	<i>Itgal, Csf2</i>
Rechazo de aloinjerto	<i>Il12b</i>
Artritis reumatoide	<i>Tlr2, Ltb, Atp6v0d2, Ccl12, Ccl5, Cd28, Acp5, Ccl2, Tnfsf13b</i>
Citotoxicidad mediada por células NK	<i>Pik3r1, Shc1, Cd48, Klrk1, Prkcb, Syk, Fcgr4, Icam1, Kras, Vav3, Calr, Ptpn6, Pik3cd, Vav1, Hcst, Plcg2, Rac2, Ulbp1, Tyrobp, Fcer1g, Ptk2b, Lcp2, Pak1, Sh2d1a, Map2k1, Sh3bp2, Ifnar2, Klrc2, Cd244a, Tnfsf10, Fyn, Bid, Icam2, Fas, Raf1, Nras, Casp3, Ifnar1</i>

## 7.2 Apéndice B. Publicaciones relacionadas



CENTRO DE BIOLOGÍA MOLECULAR SEVERO OCHOA

**Manuel Fresno Escudero**, catedrático de microbiología de la Universidad Autónoma de Madrid y jefe de línea del Centro de Biología Molecular Severo Ochoa (CSIC-UAM) y **Núria Gironès Pujol**, profesora titular de microbiología de la Universidad Autónoma de Madrid,

#### CERTIFICAN

Que han dirigido la tesis doctoral titulada: Análisis de la diversidad genómica y transcriptómica de *Trypanosoma cruzi* y su relación con la enfermedad de Chagas, realizada por D. **Francisco Callejas Hernández**.

En Madrid a tres de octubre del 2019.

Manuel Fresno Escudero

Núria Gironès Pujol